Aalborg Universitet

AALBORG UNIVERSITY

An integrative computational modelling of music structure apprehension

Lartillot, Olivier

Published in:
Proceedings, ICMPC-APSCOM 2014 Joint Conference

Publication date:
2014

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Lartillot, O. (2014). An integrative computational modelling of music structure apprehension. In M. Kyoung Song (Ed.), *Proceedings, ICMPC-APSCOM 2014 Joint Conference: 13th International Conference on Music Perception and Cognition* (pp. 80-86). Yonsei University. http://www.icmpc-apscom.org/file/icmpc-apscom_2014_(Proceedings).pdf

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: July 04, 2025

Proceedings

Edited by Moo Kyoung Song

THE 13TH INTERNATIONAL CONFERENCE ON MUSIC PERCEPTION AND COGNITION AND THE 5TH CONFERENCE FOR THE ASIAN-PACIFIC SOCIETY FOR COGNITIVE SCIENCES OF MUSIC

ICMPC-APSCOM 2014

Joint Conference

THE 13TH INTERNATIONAL CONFERENCE ON MUSIC PERCEPTION AND COGNITION AND THE 5TH CONFERENCE FOR THE ASIAN-PACIFIC SOCIETY FOR COGNITIVE SCIENCES OF MUSIC

August 4–8, 2014
Yonsei University
Seoul, South Korea

13th International Conference for Music Perception and Cognition

5th Conference of Asia-Pacific Society for the Cognitive Sciences of Music

ICMPC-APSCOM 2014

Joint Conference

Proceedings

Edited by **Moo Kyoung Song**

College of Music, **Yonsei University,**

Seoul, South Korea

August 4-8, 2014

Proceedings of the ICMPC-APSCOM 2014 Joint Conference:

13th Biennial International Conference for Music Perception and Cognition
5th Triennial Conference of the Asia Pacific Society for the Cognitive Sciences of Music

Edited by
Moo Kyoung Song

Cover design and Printed by Novias
e-novias.co.kr

Published by the College of Music, Yonsei University
www.yonsei.ac.kr

13th ICMPC-5th APSCOM Joint Conference
www.icmpc-apscom.org
- ICMPC WebPage : www.icmpc.org

The Proceedings are also available online at the
conference's website : www.icmpc-apscom.org

Copyright Moo Kyoung Song

Copyright © 2014 ICMPC13-APSCOM5 All rights reserved.

© Copyright of the content of an individual abstract is held by the first named (primary) author of the abstract.

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government.
(이 발표논문집은 2014년도 정부재원(교육부)으로 한국연구재단의 지원을 받아 발간되었음)

Welcoming Address by the Conference Chair



Dear delegates,

With all my hearts, I welcome all participants here in Seoul, Korea for the joint meeting of the 13th International Conference on Music Perception and Cognition (ICMPC) and the 5th Triennial Conference of the Asia-Pacific Society for the Cognitive Scienes of Music (APSCOM). The conference is organized by the College of Music at Yonsei University, Korean Society for Music Theory (KSMT), and Korean Society of Music Perception and Cognition (KSMPC).

This year's joint conference is the third joint international meeting of ICMPC and APSCOM. Two years ago at Thessaloniki, Austria was generous enough to hand over for us to host the conference in Seoul. Since then, we have sincerely prepared the conference for a great success. We are grateful to APSCOM and above all to Suk Won Yi and SunHee Chang for their steady support and encouragement throughout this long preparatory period. Many thanks are due to National Research Foundation of Korea, DaehanSteel, and Korea Tourism Organization for assisting us in securing a credible financial environment for the conference. We would also like to express our gratitude to the members of the international ICMPC - APSCOM 2014 Conference Advisory Board for trusting us despite the sudden change of our organizing committee.

The conference brings together leading researchers from different areas of music cognition and perception. A large number of papers from a broad range of disciplines such as psychology, psychophysics, philosophy, neuroscience, artificial intelligence, psychoacoustics, linguistics, music theory, anthropology, cognitive science, and music education were submitted and accepted by a strict blind review. Out of 332 submissions, 201 papers were selected for spoken presentation and 93 for poster presentation. Additionally, four keynote addresses will be presented in plenary sessions by four internationally distinguished colleagues. The two SEMPRE - ICMPC13 Young Researcher Award winners for this year will also present their work in plenary sessions on Thursday and Friday morning.

We also tried to provide an interesting and diverse social program. Apart from the welcome reception and banquet, a special concert for participants and a Seoul tour which will expose you to diverse Korean culture through a variety of activities will be offered on Tuesday and Wednesday. We would like to draw your attention to the special concert on Tuesday evening that features two meaningful performance, a clarinet ensemble consisting of hearing impaired adolescents and a Gayaguem solo performance by Professor YI Jiyoung. On the middle day of the conference, Wednesday, a Seoul Tour awaits you.

Closing this opening comment, I would like to thank all the members in the organizing committee, my colleagues in the College of Music and my collaborators of KSMT and KSMPC. I want to thank especially Dr. Yeajin Kim, So Yung Ahn and Eun Ju Sim, for their invaluable devoted help in various stages of this organization. Finally, a warm thanks to all of you for coming to Seoul located in the far-east Asia. We are confident that this conference will be a most rewarding and memorable experience for all.



Moo Kyoung Song,

Ph.D. in Music Theory Associate Professor, College of Music at Yonsei University & President of KSMT and KSMPC

Welcoming Address by APSCOM President



It is our great pleasure to welcome so many people to a historical place in the Asia-Pacific area, which is defined in this case as the region including and surrounding Australia, China, Japan, and South Korea. These countries have very different cultures, but we enjoyed very much working together to make it possible to hold this meeting in South Korea. This was a good opportunity for us to realize that scientific research has no borders--especially when it is related to our common language: music. Please, share this feeling with us. Korea has many attractive places, and we will meet one another both in and outside the conference venue. We will be like friends for one week. I very much hope that there will be real friendship here and there after the conference.
I am very grateful to our Korean colleagues (including my predecessor, Sun Hee Chang) who worked very hard to prepare for the present meeting here at Yonsei University.

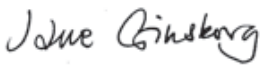


Yoshitaka Nakajima
President of Asia-Pacific Society for the Cognitive Sciences of Music

Welcoming Address by ESCOM President



On behalf of the European Society for the Cognitive Sciences of Music, I am delighted to welcome delegates to the 13th International Conference on Music Perception and Cognition and the 5th Conference of the Asia-Pacific Society for the Cognitive Sciences of Music. Whether you have come to South Korea for the first time, like me, or are making a return visit, it is a very special pleasure to be hosted by our colleagues at Yonsei University in Seoul, whom I should like to thank for their hospitality. I know from taking part in the deliberations of the ICMPC Executive Committee that every effort has been made to ensure that the conference will be successful, and, from the excellent submissions considered by the Scientific Committee, that it will be highly stimulating.
The biennial International Conference provides a wonderful opportunity for all of us to come together and focus, for five days, on the huge variety of recent research that is being undertaken in our different disciplines. We can look forward to learning about new approaches, methods and findings, and being inspired by new ideas; forming new international partnerships and consolidating existing collaborations and friendships. I wish you a productive conference and an enjoyable stay in Seoul.



Jane Ginsborg
President of the European Society for the Cognitive Sciences of Music

Welcoming Address by SMPC President



Dear ICMPC 2014 Delegates, On behalf of the Society for Music Perception and Cognition, I am delighted to welcome you to the 13th biennial International Conference on Music Perception and Cognition. As our field grows steadily in size and stature, conferences like ICMPC are essential to the dissemination and coordination of research, and I wish to express my sincere thanks to the organizers of this year's meeting in Seoul. SMPC is historically a North American society, however we have members around the globe, including Europe, Asia and Australia, and we are excited to participate in this year's meeting. I am also pleased to announce that SMPC's next biennial meeting will be held on August 1-5, 2015, and hosted at Vanderbilt University with Elizabeth Dykens as conference chair, and Reyna Gordon as conference co-chair. Additional details will be available soon. In addition, the next ICMPC will be held in North America, and I plan to announce dates and venue at this year's meeting. To learn more about SMPC, please visit our website, www.musicperception.org, where you can read recent news, learn about upcoming events, find information about our laboratories and graduate programs, and watch videos and podcasts that feature the latest research in music perception and cognition.
Best wishes for a stimulating and productive meeting,



Ed Large
President of Society for Music Perception and Cognition

ICMPC13-APSCOM5 CONFERENCE TEAM

Conference Chair | Moo Kyoung Song

Organizing Committee

Kyungil Kim *(co-chair)*
Suk Won Yi *(program coordinator)*
Kyung Myun Lee *(secretary general)*
Yeajin Kim *(senior officer)*
Hye-yoon Chung *(officer)*
So-Yung Ahn *(treasurer)*

Conference Secretaries

Sol Lim
Eun Ju SIM

Program/Abstract Book Coordinators

Yi Eun Chung
You Jin Kim

Editorial Assistants

Jeong-Mi Park
Woo-Ah Min
Donggyu Yoo

Conference Helper

Plan De Communication

ICMPC13-APSCOM5 Volunteers

Eun Ju SIM (coordinator)
Jung Yun Lee
Koh Yunhwa
Sae Hee Park
Daehye Bae
You Jin Kim
Goeun Lee
Yeoeun Lim
Lee Yeon Woo
Han Gyeul Chung
Woo-Ah Min
Donggyu Yoo
Chung JooHyun
Kim Hee Jung
Kim SangHon
Yunyoung Choi

Organizing Committee

Moo Kyoung Song	*Chair, Yonsei University, College of Music*
Kyungil Kim	*Co-chair, Ajou University, Department of Psychology*
Suk Won Yi	*Seoul National University, Department of Musicology*
So-Yung Ahn	*Hanyang University, College of Music*
Hye-yoon Chung	*Myongji University, Division of Music*
Yeajin Kim	*Yonsei University, College of Music*
Kyung Myun Lee	*Seoul National University, Graduate School of Convergence Science and Technology*

Program Committee

Robert Gjerdingen	*Northwestern University*
Petr Janata	*University of California, Davies*
Youn Kim	*University of Hong Kong*
Reinhard Kopiez	*Hannover Hochschule für Musik, Theater und Medien*
Ian Quinn	*Yale University*
John Rink	*University of Cambridge*
Catherine Stevens	*University of Western Sydney*
Barbara Tillmann	*Lyon Neuroscience Research Center*
Suk Won Yi	*Seoul National University, (coordinator)*

Conference Advisory Board

Mayumi Adachi	*Chair of ICMPC 10, Hokkaido University*
Emilios Cambouropoulos	*Chair of ICMPC 12, Aristotle University of Thessaloniki*
Steven M. Demorest	*Chair of ICMPC 11, University of Washington*
Jane Ginsborg	*President of ESCOM, Royal Northern College of Music*
Etsuko Hoshino	*President of JSMPC, Ueno Gakuen University*
Edward Large	*President of SMPC, University of Connecticut*
Sarah Wilson	*President of AMPS, University of Melbourne*
Shibin Zhou	*President of CSMP, Capital Normal University*

Scientific Advisory Board

Mayumi Adachi	*Hokkaido University, Japan*
Anna Rita Addessi	*University of Bologna, Italy*
So-Yung Ahn	*Hanyang University, Korea*
Rita Aiello	*New York University, USA*
Rytis Ambrazevičius	*Kaunas University of Technology, Lithuania*
Richard Ashley	*Northwestern University, USA*
Warren Brodsky	*Ben-Gurion University of the Negev, Israel*
Emilos Cambouropoulos	*Aristotle University of Thessaloniki, Greece*
Hyun Ju Chong	*Ewha Womans University, Korea*
Chun Kee Chung	*Seoul National University, Korea*
Hye-yoon Chung	*Myongji University, Korea*
Moonhyuk Chung	*Ewha Womans University, Korea*
Eugenia Costa-Giomi	*University of Texas, Austin, USA*
Sarah Creel	*University of California, San Diego, USA*
Ian Cross	*University of Cambridge, UK*
W. Jay Dowling	*University of Texas, Dallas, USA*
Tuomas Eerola	*University of Jyväskylä, Finland*
Morwaread Farbood	*New York University, USA*
Takako Fujioka	*Stanford University, USA*
Jane Ginsborg	*Royal Northern College of Music, UK*
Robert Gjerdingen	*Northwestern University, USA*
Werner Goebl	*University of Music and Performing Arts, Austria*
Jessica Grahn	*Western University, Canada*
Anderea Halpern	*Bucknell University, USA*
Stephen Handel	*University of Tennessee, USA*
Etsuko Hoshino	*Ueno Gakuen University, Japan*
Erkki Huovinen	*University of Minnesota, USA*
David Huron	*Ohio State University, USA*
Cong Jiang	*Capital Normal University, China*
Roger Kendall	*University of California, Los Angeles, USA*
Jinho Kim	*Andong National University, Korea*
Soo Ji Kim	*Ewha Womans University, Korea*
Yeajin Kim	*Yonsei University, Korea*
Youn Kim	*The University of Hong Kong, Hong Kong*
Oyeon Kwon	*Yonsei University, Korea*
Alexandra Lamont	*Keele University, UK*
Eleni Lapidaki	*Aristotle University of Thessaloniki, Greece*
Kyogu Lee	*Seoul National University, Korea*
Kyung Myun Lee	*Seoul National University, Korea*
Scott Lipscomb	*University of Minnesota, USA*
Steven Livingstone	*Ryerson University, Canada*
Raymond MacDonald	*Glasgow Caledonian University, UK*
Elizabeth Margulis	*University of Arkansas, USA*
Stephen McAdams	*McGill University, Canada*
Yoshitaka Nakajima	*Kyushu University, Japan*
Takayuki Nakata	*Future University, Japan*
Seung-Ha Oh	*Seoul National University, Korea*
Marta Olivetti	*Sapienza University of Rome, Italy*
Georgios Papadelis	*Aristotle University of Thessaloniki, Greece*
Nikki Rickard	*Monash University, Australia*
Martina Rieger	*University for Health Sciences, Medical Informatics and Technology, Australia*
Jaan Ross	*Estonian Academy of Music and Theatre, Estonia*
Frank Russo	*Ryerson University, Canada*
Emery Schubert	*University of New South Wales, Australia*
Uwe Seifert	*University of Cologne, Germany*
Catherine Stevens	*University of Western Sydney, Australia*
David Temperley	*Eastman School of Music, USA*
Petri Toiviainen	*University of Jyväskylä, Finland*
Minoru Tsuzaki	*Kyoto City University of Arts, Japan*
Oliver Vitouch	*University of Klagenfurt, Austria*
Sarah Wilson	*The University of Melbourne, Australia*
Suk Won Yi	*Seoul National University, Korea*

sempre:

Society for Education, Music and Psychology Research

SEMPRE AWARDS

The Society for Education, Music and Psychology Research (SEMPRE) http://www.sempre.org.uk/ kindly offers a number of awards to researchers attending this year's ICMPC conference

ICMPC13 Young Researcher Award (sponsored by SEMPRE)

The ICMPC 13-APSCOM 5 Young Researcher Award (YRA), sponsored by Society for Education, Music and Psychology Research (SEMPRE), is awarded to young researchers that submit a high quality research paper and demonstrate the potential to be a leading researcher in the field of Music Perception and Cognition.

This year's YRA was decided by the eight members of ICMPC-APSCOM 2014 Program Committee, consisting of :

Robert Gjerdingen *(Northwestern University, USA)*
Petr Janata *(University of California, Davis, USA)*
Youn Kim *(The University of Hong Kong, Hong Kong)*
Reinhard Kopiez *(Hannover Hochschule für Musik, Theater und Medien, Germany)*
John Rink *(University of Cambridge, UK)*
Catherine Stevens *(University of Western Sydney, Australia)*
Barbara Tillmann *(Lyon Neuroscience Research Center, France)*
Suk Won Yi *(Seoul National University, Korea), Coordinator.*

The referees examined carefully all applications, and decided this year's YRA prize to be shared by the following two researchers:

Lauren Hadley *(University of Edinburgh, UK)*
Sook Young Won *(Stanford University, USA)*

The selection process consisted of the following steps: Nineteen applications were received and each submission was evaluated by three referees. Then, applications were shortlisted; the papers of four finalists were carefully examined by seven members of the Program Committee in terms of their overall quality and originality, and, in terms of meeting all the criteria described on the conference webpage, delivered their final decision.

Apart from receiving a money prize (1,500 US dollars each), the two YRA winners will present their work in special plenary sessions on August 5 and 7, 2014.
The organizers of ICMPC13-APSCOM 5 would like to thank SEMPRE for generous support and congratulate wholeheartedly the winners for their success.

Travel Award (sponsored by SEMPRE)

The Travel Awards are awarded by SEMPRE to financially assist ICMPC participants on the basis of merit and need. This year, a total of 13,900 US dollars has been awarded to the following participants: Bruno ALCALDE, Rasmus Arnling BååTH, Jenine BROWN, Nicola DI STEFANO, Alison FARLEY, Jesper HOHAGEN, Kelly JAKUBOWSKI, Karl LERUD, Xiaoluan LIU, Xuejing LU, Megha MAKAM, Stephanie Audrey MCCULLOUGH, Ming Wai PANG, Johan PAULIN, Olivia M. PODOLAK, Sarah Anne SAUVE, Pat SAVAGE, Kai SIEDENBURG, Li-Ching WANG, Michael WEISS, and Suki YIU. The organizers of ICMPC 13-APSCOM 5 would like to thank SEMPRE for generous support and congratulate wholeheartedly the winners for their success.

Conference Programme at a Glance

<table>
<tr><th></th><th>Monday
4 August</th><th>Tuesday
5 August</th><th>Wednesday
6 August</th><th>Thursday
7 August</th><th>Friday
8 August</th></tr>
<tr><td>8:30-9:00</td><td rowspan="2">Registration</td><td rowspan="4">Session 3</td><td rowspan="20">Excursions</td><td rowspan="4">Session 6</td><td rowspan="4">Workshops
&Session 9</td></tr>
<tr><td>9:00-9:30</td></tr>
<tr><td>9:30-10:00</td><td rowspan="5">Welcome
&
Keynote 1</td></tr>
<tr><td>10:00-10:30</td></tr>
<tr><td>10:30-11:00</td><td>Break</td><td>Break</td><td>Break</td></tr>
<tr><td>11:00-11:30</td><td rowspan="3">Keynote 2</td><td rowspan="3">Keynote 3 & YRA1</td><td rowspan="3">Session 10</td></tr>
<tr><td>11:30-12:00</td></tr>
<tr><td>12:00-12:30</td><td rowspan="2">Opening
Reception</td></tr>
<tr><td>12:30-13:00</td><td rowspan="3">Lunch</td><td rowspan="3">Lunch</td><td rowspan="3">Lunch</td></tr>
<tr><td>13:00-13:30</td><td rowspan="2">Poster Session 1</td></tr>
<tr><td>13:30-14:00</td></tr>
<tr><td>14:00-14:30</td><td rowspan="4">Session 1</td><td rowspan="4">Symposium1
&
Session 4</td><td rowspan="4">Symposium2 &
Session 7</td><td rowspan="4">Session 11</td></tr>
<tr><td>14:30-15:00</td></tr>
<tr><td>15:00-15:30</td></tr>
<tr><td>15:30-16:00</td></tr>
<tr><td>16:00-16:30</td><td>Break</td><td>Break (16:00-16:15)</td><td>Break (16:00-16:15)</td><td>Break</td></tr>
<tr><td>16:30-17:00</td><td rowspan="4">Session 2</td><td rowspan="3">Poster Session 2
&
Session 5</td><td>Poster Session 3</td><td rowspan="3">Keynote 4 & YRA2</td></tr>
<tr><td>17:00-17:30</td><td rowspan="3">Session 8</td></tr>
<tr><td>17:30-18:00</td></tr>
<tr><td>18:00-18:30</td><td rowspan="3">ICMPC
Assembly</td></tr>
<tr><td>18:30-19:00</td><td rowspan="3">Program
Committee</td><td rowspan="3">Banquet</td></tr>
<tr><td>19:00-19:30</td></tr>
<tr><td>19:30-20:00</td></tr>
<tr><td>20:00-20:30</td><td rowspan="2">APSCOM
Assembly</td><td></td></tr>
<tr><td>20:30-21:00</td><td>Concert</td></tr>
</table>

Programme

ICMPC 13-APSCOM 5

August 4-8, 2014
Yonsei University
Seoul, South Korea



ICMPC
APSCOM
2014 SEOUL

Program Committee

Robert Gjerdingen (Northwestern University, USA)
Petr Janata (University of California, Davis, USA)
Youn Kim (The University of Hong Kong, Hong Kong)
Reinhard Kopiez (Hannover Hochschule für Musik, Theater und Medien, Germany)
Ian Quinn (Yale University, USA)
John Rink (University of Cambridge, UK)
Catherine Stevens (University of Western Sydney, Australia)
Barbara Tillmann (Lyon Neuroscience Research Center, France)
Suk Won Yi (Seoul National University, Korea), Coordinator

Monday 4 August			
08:30-09:30	**Registration - Baekyang Building Lobby**		
09:30-09:50	**Opening Ceremony** (Chaired by Kyung Myun Lee, Secretary general), 1. Congratulatory Remark- Kap-Young Jeong, President of Yonsei University 2. Welcoming Address-Moo Kyoung Song, Organizing Chair of the Conference 3. Welcoming Addresses from the presidents of 3 primary regional conferences ➢ Yoshitaka Nakajima, APSCOM President ➢ Jane Ginsborg, ESCOM President ➢ Ed Large, SMPC President		
09:50-12:00	**Welcome & Keynote 1 – SEOUL** Robert GJERDINGEN, Chair Sukjae LEE: *Esse est percipi*: Some Early Modern Perspectives on the Fundamental Mind-Dependency of Objects		
12:00-13:00	**Opening Reception - Baekyang Building Lobby**		
	Poster Session 1 - Baekyang Building Lobby		
13:00-14:00	[1-01] Gerard REMIJN, Yushiro TSUBAKI, Kazuo UEDA, Yoshitaka NAKAJIMA: Auditory Reorganization of Gliding Tones in Different Frequency Ranges [1-02] Hidetaka IMAMURA, Sungyoung KIM, Atsushi MARUI, Toru KAMEKAWA, Richard KING, Wieslaw WOSZCZYK: Discovering Perceptual Characteristics of Multichannel Music with Height Ambiences for Japanese Listeners [1-03] Satoshi OKAZAKI, Makoto ICHIKAWA: Perceptual Fusion and Simultaneity for Auditory Stimuli [1-04] Hyun Sung NAM: The Effect of Audio Panning on the Perception of Movement in Film [1-05] Iku NEMOTO, Satoshi UENO: Ambiguous [1-06] Christopher WHITE, Ian QUINN: Compiling and Processing the Yale-Classical Archives Corpus [1-07] Dale MISENHELTER: Examining the Anticipatory: Meta-Analysis of Response to Structural Functions [1-08] Andreu BALLÚS, Eric ARNAU, Oriol NIETO, Frederic FONT, Alba G. TORRENTS: Embodying Theoretical Research in Music Cognition: Four Proposals for Theory-Driven Experimentati	[1-09] Mizuki YAMASAKI, Masanobu MIURA: Use of Eigenmusic to Arrange Music Excerpts [1-10] Sun Hee LEE: Error Culture in Instrument Lessons in South Korean Music Schools [1-11] Stephanie MCCULLOUGH, Elizabeth MARGULIS: The Effect of Motor Involvement and Melody Truncation on Involuntary Musical Imagery [1-12] Rhimmon SIMCHY-GROSS, Elizabeth MARGULIS: Modulating Attention in Multi-Part Music: Is Simple Consonance More Helpful than Compound Dissonance? [1-13] Philip FINE, Lucy KIRBY: The Effect of Music Tempo and Background Noise on Adults' Reading Comprehension and Reading Speed [1-14] Georgina FLORIDOU, Lauren STEWART, Victoria WILLIAMSON, Daniel MÜLLENSIEFEN: Development and Validation of the Involuntary Musical Imagery Scale (IMIS) [1-15] Nicolas FARRUGIA, Kelly JAKUBOWSKI, Robert CARLYON, Rhodri CUSACK, Lauren STEWART: Reported Pleasantness of Involuntary Musical Imagery Correlates with Gray Matter Density in Emotion-Related Areas	[1-16] Stefania PILERI, Cristina DI BERNARDO, Alan D BADDELEY, Graham J HITCH, Daniel MULLENSIEFEN, Victoria J WILLIAMSON: Mapping Musicians' Memory [1-17] Yoko OGAWA: Occurrence and Process of the Performer's Anxiety for Opera and Piano Concert by University Students [1-18] Justin BLACK, Elizabeth MARGULIS: Contextual Variability in Affective Responses to Music [1-19] Madeline HUBERTH, Sarah HAWKINS, Ian CROSS: Arousal Facilitates Recognition Memory for Unfamiliar Music [1-20] Chelsea DOUGLAS, Stephen MCADAMS: Perceived Affect of Musical Instrument Timbres [1-21] Malinda MCPHERSON, Monica LOPEZ-GONZALEZ, Summer RANKIN, Charles LIMB: Musical Features of Spontaneous Improvisation Associated with Emotional Cues [1-22] Junko MATSUMOTO: Psychological Effects of 'Hitokara' Singing on Mood

Room	SEOUL	KYOTO	MONTREAL	LIÈGE	LOS ANGELES
Session 1	**[1A] Music and Evolution**	**[1B] Music in Everyday Life**	**[1C] Music and Movement 1**	**[1D] Aesthetic Perception and Emotions**	**[1E] Acoustics and Psychoacoustics**
Chair	Helen FARRELL	Wilfried GRUHN	Petri TOIVIAINEN	Jin Hyun KIM	Permagnus LINDBORG
14:00-14:30	Kathryn DAVIES : Researching the Origins of Music with Methods from First Language Acquisition Research	Naomi ZIV: The Effect of Positive Background Music on Compliance	Li-Ching WANG: Do Listener's Body Movements Influence Their Perception of Rhythm?	Paul ELVERS, Wolfgang FUHRMANN, Timo FISCHINGER: Musical Preferences of Musicology Students	Takuya KISHIDA, Yoshitaka NAKAJIMA, Kazuo UEDA, Gerard B. REMJIN: Perceptual Roles of Power Fluctuation Factors in Speech
14:30-15:00	Piotr PODLIPNIAK: Pitch Centricity as an Evolutionary Innovation of Vocal Communication	Amanda KRAUSE, Adrian NORTH: Contextualized Listening: Playlists for Everyday Situations	Kyung Myun LEE: Does Watching Body Movement Affect Rhythm Perception?	Hans NEUHOFF, Rainer POLAK, Timo FISCHINGER: Perception of Metric Timing Patterns in Malian Jembe-Music	Sven-Amin LEMBKE, Scott LEVINE, Martha DE FRANCISCO, Stephen MCADAMS: The Influence of Performer Roles on Timbre Blending Between Bassoon and Horn Players
15:00-15:30	Peter KELLER, Rasmus KOENIG: Social Communicative Functions of the Singers' Formant in Choral Music	Hauke EGERMANN, Daniel KAHLHOEFER, Christof M. SCHULTZ, Marten SEEDORF: New Media Use and Music Appreciation: Connections between Contemporary Access to Music and Its Perceived Value	Birgitta BURGER, Geoff LUCK, Marc THOMPSON, Suvi SAARIKALLIO, Petri TOIVIAINEN: I Like to Move It: How Music-Induced Movement Relates to Music Preference, Familiarity, and Movability	Suvi SAARIKALLIO, Antti YRTTI, Minna HUOTILAINEN: Emotional Expression through Music in 3-5-Year-Olds	Jessica GILLARD, Michael SCHUTZ: A Quantitative Analysis of Auditory Stimuli: New Results and Insights
15:30-16:00	Yuko HATTORI, Masaki TOMONAGA, Tetsuro MATSUZAWA: Spontaneous Auditory-Motor Entrainment during Self-Paced Tapping in Chimpanzees and Humans		Esther COOREVITS, Dirk MOELANTS: The Interpretation of Tempo in Baroque Dance and Music	William RANDALL, Nikki RICKARD: A Model of Personal Music Listening: Emotional Outcomes Within Contexts and between Listeners	Michael OEHLER, Christoph REUTER, Isabella CZEDIK-EYSENBERG, Niklas HILL, Michael ZIETHEN: Psychoacoustical and Emotional Cues in German Popular Music
16:00-16:30	**Break**				

Room	SEOUL	KYOTO	MONTREAL	LIÈGE	LOS ANGELES
Session 2	**[2A] Musical Timbre**	**[2B] Listening as a Creative Process**	**[2C] Musicking**	**[2D] Music and Movement 2**	**[2E] When Music Delivers Pleasure**
Chair	Stephen MCADAMS	Hasan Gürkan TEKMAN	Peter KELLER	Birgitta BURGER	Elizabeth MARGULIS
16:30-17:00	Morwaread FARBOOD, Khen PRICE: Timbral Features Contributing to Perceived Auditory and Musical Tension	Frieder STOLZENBURG: Harmony Perception by Periodicity Detection	Andrew GOLDMAN: Delayed Auditory Feedback Disrupts Improvised Piano Playing More Than Memorized Playing: Towards a Cognitive Characterization of Improvisation	Mari Romarheim HAUGEN, Rolf Inge GODØY : Rhythmical Structures in Music and Body Movement in Samba Performance	Ji Yeon LEE: "Delay" in Climax: Contextualization, Analysis, and Implication
17:00-17:30	Kai SIEDENBURG, Stephen MCADAMS: Context Effects in the Cognitive Sequencing of Musical Timbre	Nicola DI STEFANO: Consonance and Dissonance Perception in Infants: Theoretical Considerations and Empirical Proposal	Freya BAILES, Roger DEAN: Perceptions of Leadership in Duo Keyboard Improvisations	Megha MAKAM, Blair KANESHIRO, Jonathan BERGER: Capoeira Interaction as a Model of Expectation Formulation and Violation in Real-Time Improvised Performance	Chen-Gia TSAI, Chung-Ping, Szu-Pei YU: When the Musical Reward Arrives: Listeners' Physiological Responses to the Theme Entrance/Recurrence in the Verse-chorus Form and the Sonata Form
17:30-18:00	Permagnus LINDBORG: Congruency of Colour Associations to Urban Soundscapes, Restaurants, and Electroacoustic Soundscape Compositions	Edgar CARDOSO: Creating Music That Moves: Creative Drama and Movement as Tools for Improvisation in Music	Massimo AVANTAGGIATO: Donatoni: A Critical Re-Appraisal of Quartetto III	Peta MILLS, Peter KELLER: The Relationship between Temporal Anticipation and Adaptation during Sensorimotor Synchronisation	Johan PAULIN, Guy MADISON: Music with Varying Degrees of Complexity Are Rated Differently on Liking and Complexity by Men Compared to Women
18:00-18:30		Haruka SHODA, Mayumi ADACHI: Music is Communicated Visually: Modeling Performer-to-Audience Communication in Live Piano Performance	Nathaniel CONDIT-SCHULTZ: Rhyme, Reason, and Rhythm: Elision, Enjambment and Entropyin the Phrasing and Rhyme Schemes of Rap	Laura BISHOP, Werner GOEBL: Effects of Musical Expertise on Audiovisual Integration: Instrument-Specific or Generalisable?	
20:00-	**APSCOM General Assembly – MONTREAL**				

Tuesday 5 August					
Room	**SEOUL**	**KYOTO**	**MONTREAL**	**LIÈGE**	**LOS ANGELES**
Session 3	**[3A] Cognitive Neuroscience of Music 1**	**[3B] Rhythm and Timing**	**[3C] Musical Structure**	**[3D] Social Psychology of Music**	**[3E] Power of Music**
Chair	Frank RUSSO	Yoshitaka NAKAJIMA	Freya BAILES	Naomi ZIV	Joshua ALBRECHT
08:30-9:00	Anna-Katharina Raphaela BAUER, Gunter KREUTZ, Christoph HERRMANN: Electrophysiological Correlates of Musical Tempo Preferences	Alison FARLEY: An Evaluation of Musicians' Internal Pulse and Rhythmic Sight-Reading	Marlies BODINGER, Erica BISESI, and Richard PARNCUTT: Listeners' Informal Vocabulary for Structure, Expression, Emotions and Associations in Piano Music	Wilfried GRUHN, Kristi KIILU, Reet RISTMÄGI, Karin TÄHT: Impact Factors on Musical Identity Formation of Professional Musicians and Music Teachers	Ki-Hong KIM, Shin-Ichiro IWAMIYA, Fumiya MORI: Effectiveness of Sound Effects and Music to Induce Laugh in Comical Entertainment Television Show
09:00-09:30	Sujin HONG, Neil ROBERTS, Alexa MORCOM, Katie OVERY: The Role of the Inferior Frontal Gyrus in Beat and Metre Perception	Rasmus BÅÅTH, Kristin Osk INGVARSDOTTIR: Subjective Rhythmization: A Replication and a Reassessment	Elizabeth MARGULIS: Verbatim Repetition and Listening Musically	Zhuo ZHANG: Music "Violence": Hurt behind the Sounds	Mark REYBROUCK: Music and Its Inductive Power: A Psychobiological and Evolutionary Approach to Musical Meaning
09:30-10:00	Emi HASUO, Kazuo UEDA, Takuya KISHIDA, Haruna FUJIHIRA, Satoshi MORIMOTO, Gerard B. REMIJN, Kimio SHIRAISHI, Shozo TOBIMATSU, Yoshitaka NAKAJIMA: Duration Perception of Filled and Empty Intervals: A Study with Magnitude Estimation and Electroencephalography	Fiona MANNING, Michael SCHUTZ: Tapping through Longer Periods of Silence Improves Perceived Timing	Bruno ALCALDE: Frequency as Determinant of Function and Structure: A Usage-Based Approach to JoÃo Gilberto's Guitar Accompaniment in Bossa Nova	Tommi HIMBERG, Neta SPIRO: Interactionism and Music: Measuring Entrainment and Influence in Dyads and Groups	Richard VON GEORGI, Kai Stefan LOTHWESEN: Use of Music for Emotion Modulation:Testing the Validity of the Iaam-Scales within Different Groups of Musical Preference Indicates Different Qualitative Music Preference Types
10:00-10:30	Iballa BURUNAT, Elvira BRATTICO, Tuomas PUOLIVÄLI, Tapani RISTANIEMI, Mikko SAMS, Petri TOIVIAINEN: Prominent Interhemispheric Functional Symmetry in Musicians	Jason MUSIL, Daniel MÜLLENSIEFEN: Optimising a Test of the Ability to Tap in Time with Musical Beat	Olivier LARTILLOT: An Integrative Computational Modelling of Music Structure Apprehension	Hasan Gürkan TEKMAN: All the Lovers are Tenors: Evolutionary Psychology and Queer in Verdi	Hye Young PARK: Difference in Affective Responses to Music between Congenitally Blind Sighted Adults
10:30-11:00	**Break**				
11:00-12:30	**Keynote 2 - SEOUL** Reinhard KOPIEZ, Chair Catherine STEVENS: Learning and Memory Processes in Music and Dance				
12:30-14:00	**LUNCH**				

Room	SEOUL	KYOTO	MONTREAL	LIÈGE	LOS ANGELES
Session 4	**[4A] Symposium 1**	**[4B] Music and Movement 3**	**[4C] Music and Health**	**[4D] Music Performance, Expression and Objective Analysis**	**[4E] Music and Emotions**
Chair	Daniel SHANAHAN	Kyung Myun LEE	Hyun Ju CHONG	Emery SCHUBERT	Mark REYBROUCK
14:00-14:30	***What Empirical Musicology Offers Emotion Studies*** Discussant: Richard ASHLEY Elizabeth Hellmuth MARGULIS: Interest as an Emotional Response to Music Diana MARTINEZ, Lorraine CHUEN, Michael SCHUTZ: From score to stage: Exploring the relationship between compositional structure and emotional response Daniel SHANAHAN: Using Sentiment Analysis and Probabilistic Topic Modeling to Examine the Tessitura: Sociability Relationship of Operatic Characters David HURON: Integrating Psychological, Musicological and Ethological Perspectives on Music and Emotion	Marek FRANEK, Leon VAN NOORDEN, Lukas REZNY: The Walking with Music in an Urban Environment	Jill MORGAN, Raymond MACDONALD, Paul FLOWERS, Dorothy MIELL: Active Ageing: The Symbiosis of Music and Health in Third Age Relationships	Ricardo GOLDEMBERG, Philip FINE: Evaluation of a Bottom-up Model of Musical Sight Singing: The Case of Modus Novus	Finn UPHAM: Reframing Emotions and Music
14:30-15:00		Jin Hyun KIM: Kinaesthetic Dimensions in the Aesthetic Perception and Experience of Music	Gunter KREUTZ: Does Singing Facilitate Social Bonding?	Masanobu MIURA: Eigenperformance: Method to Describe the Proficiency of Piano Performance	Hyun Ju CHONG, Ga Eul YOO, Soo Ji KIM: Affective Responses to Complexity in Harmonic Progression between Music and Nonmusic Majors
15:00-15:30		David MOSHER: Temporal Structures Embodied through Dance-Interactions of Perception, Cognition, and Attempted Escape from the Periodic Temporal Perception Window	Jenny GROARKE, Michael HOGAN: A Qualitative Investigation of Developmental Differences in the Functions of Music Listening and Its Role in Everyday Well-Being	Erica BISESI, Richard PARNCUTT, Sylvain CARON, Caroline TRAUBE: The Immanent Emotion of a Musical Score: An Exploratory Study	Joshua ALBRECHT: Naming the Abstract: Building a Repertoire-Appropriate Taxonomy of Affective Expression
15:30-16:00		Youn KIM: “In Search of Lost Time”: Connections between Theories of Hearing and Performance Discourse	Clorinda PANEBIANCO-WARRENS: The Role of Music in the Experience of Flow in Ballet Dancers	Johanna DEVANEY: Digitization and Analysis of Seashore's Historical Music Performance Scores	Edith VAN DYCK, Pieter VANSTEENKISTE, Matthieu LENOIR, Micheline LESAFFRE , Marc LEMAN: Emotion Recognition from Music-Induced Movement
16:00-16:15	**Break**				

	Poster Session 2 - Baekyang Building Lobby		
16:15-17:00	[2-01] Neta SPIRO, Tommi HIMBERG: Improvisation and Change in Videos of 1-To-1 Music Therapy Sessions with Children with Autism Spectrum Disorders: A Case Example [2-02] Rumi HIRAGA, Kazuhiko OTSUKA, Shu MATSUDA, Keitaro TAKAHASHI: Using Music Visualization Systems with Hearing-Impaired Children [2-03] Jesper HOHAGEN: Well-Being and Quality of Life: Self-Reflection of Adult Amateur Instrumental Students [2-04] Jenny GROARKE, Michael HOGAN: The Adaptive Functions of Music Listening: An Exploratory Factor Analysis [2-05] Yo-Jung HAN: Analysis of Verbal Descriptors for Timbre: Based on Conceptual Metaphor Theory [2-06] Hyekyung PARK: The Relationship between Music and Language in Terms of Derrida's Concept of Writing [2-07] Swathi SWAMINATHAN, Glenn SCHELLENBERG: Music Training and Linguistic Structure Processing in Children and Adults [2-08] Isabella CZEDIK-EYSENBERG, Christoph ANZENBACHER, Christoph REUTER, Michael OEHLER: Branch-Dependent Time- and Timbre-Related Features of Audio Logos [2-09] Christoph REUTER, Michael OEHLER, Jörg MÜHLHANS: Physiological and Acoustical Correlates of Unpleasant Sounds	[2-10] Frank LIEBSCHER: Role And Effects Of Procedural Orientation On Skill Acquisition In Deliberate Music Practice [2-11] Sachiyo KAJIKAWA, Hideo MORIUCHI: Musical, Language, And Social Development Of One-Year-Olds Participating In Music Classes [2-12] Tong XUE: Montreal Battery Of Evaluation Of Musical Abilities: Evaluation Of Chinese Adolescents' Musical Abilities [2-13] Xinyang HUO, Mayumi ADACHI, Emiko KUSANAGI, Nobuko HOSHI, Shing-Jen CHEN, Tadashi OISHI, Hitoshi TAKAMURA: Factors Influencing Japanese Preschoolers' Involvement In Musical Play At Home [2-14] Lingjing CHEN, Mayumi ADACHI, Masae TAGA, Hisanori MINAKAMI, Yasushi HANDA, Masanori KANEUCHI: An Exploratory Study On The Fetal Movement To Music [2-15] Sung Hyun HWANG: Music Cognition Into Practice Room: Implication For College Instrumental Teacher [2-16] Rosane CARDOSO DE ARAUJO: Approaches About Motivation For The Practice Of Music In The Higher Education: Two Discursive Possibilities [2-17] Junko TAKAHASHI: Harmonic Sensibility Of Elementary School Children On Creating A Sub-Melody [2-18] Yuki MITO, Hiroshi KAWAKAMI, Masanobu MIURA, Yukitaka SHINODA: Investigation of Performance Motion on Keyboard Instrument by Difference in Key	[2-19] Kai SIEDENBURG, Stephen MCADAMS, Petr JANATA: Outside the Phonological Loop? Serial Order and Similarity in Working Memory for Timbre [2-20] Liming SONG, Ming LI, Yonghong YAN: Vocal Melody Extraction Based on Bayesian Framwork [2-21] Jackson GRAVES, Andrew OXENHAM: Effects of Tonal Context on The Perception of Pitch Intervals [2-22] Rafał LAWENDOWSKI, Krzysztof BASIŃSKI: Interval and Beat-Based Strategies of Rhythm Perception in Musicians and Non-Musicians [2-23] Sota OKADA, Hiroshi KUWABARA, Masanobu MIURA: Ambiguity of Tempo Perception for J-Pop And Non J-Pop Tunes [2-24] Masaki MATSUBARA, Hiroko TERASAWA, Kjetil F. HANSEN, Rumi HIRAGA: An Inquiry into Hearing-Impaired Student's Musical Activities: How Do They Listen to the Music? [2-25] Hitoshi TOMINAGA, Teruo YAMASAKI, Yukiko UCHIDA, Yuri MIYAMOTO: Effects of Pair Ensemble on Attention and Affective Experience [2-26] Justin LONDON, Rainer POLAK: The Influence of Task and Context on Complex Rhythm Production: Evidence from Malian Drumming [2-27] Eun-Jeong LEE, Jens-Peter ROSE, Joachim WEIS A Clinical Study about the Effect of Music Therapeutic Treatments in an Oncological Rehabilitation

Room	SEOUL	KYOTO	MONTREAL	LIÈGE	LOS ANGELES
Session 5	**[5A] Tonality vs. Atonality**	**[5B] Musical Schemata and Cognition**	**[5C] Rhythm in Language, Music and Movement**	**[5D] Music and Memory 1**	
Chair	So-Yung AHN	Youn KIM	M. Paula RONCAGLIA-DENISSEN	Friedrich PLATZ	
16:30-17:00	Heewon CHUNG: The Limits of Tonality: Semitonal Modulations in Chopin's Music	Hye-Yoon CHUNG: A Cognitive Approach to Musical Gesture	Zhimin BAO, Yuko YAMASHITA, Kazuo UEDA, Yoshitaka NAKAJIMA: The Acquisition of Speech Rhythm in Chinese-,English-, and Japanese-Learning Infants	Georgina FLORIDOU, Victoria WILLIAMSON, Lauren STEWART: Capturing Earworms in the Lab: Effects of Cognitive Load on the Appearance of Involuntary Musical Imagery	
17:00-17:30	Jenine BROWN: Implicit Learning of Musical Structure in Twelve-Tone Music: What Do Novice Listeners Glean from the Musical Surface?	Robert GJERDINGEN: The Institutionalization of Apprenticeship in the Great Conservatories: A Cognitive Interpretation of a Non-Verbal Praxis	Niels Chr. HANSEN, Makiko SADAKATA, Marcus PEARCE: Non-Linear Changes in Rhythmic Variability of European Art Music: Quantitative Support for Historical Musicology	Niall KLYN, Yong Jeon CHEONG, Erin ALLEN, Udo WILL: Different Mechanisms for Rhythm Memory in Reproduction And Comparison Tasks	
17:30-18:00	Yi-Cheng Daniel WU: Spaciousness or Evenness? The Examination of IC-Space & Harmonic Density in Atonal Music	Daniel MÜLLENSIEFEN, Amit AVRON, Naoko SKIADA: The Structure of Cognitive and Musical Abilities	Soo Ji KIM, Hye Ji PARK: The Effects of Rhythmic Auditory Stimulation on Gait Parameters of Adolescents with Traumatic Brain Injury Depending on the Presence of Melody	Cecilia TAHER, René RUSCH, Stephen MCADAMS: Effects of Repetition on the Perception of Musical Texture	
18:00-19:00	**ICMPC General Assembly - SEOUL**				
20:30-22:00	**Concert - Yun Ju-Yong Recital Hall**				

Wednesday 6 August	
08:30-18:30	**Excursions**
18:30-20:00	**Program Committee**

Thursday 7 August					
Room	**SEOUL**	**KYOTO**	**MONTREAL**	**LIÈGE**	**LOS ANGELES**
Session 6	**[6A] Cognitive Neuroscience of Music 2**	**[6B] Multimodal Perception and Performance**	**[6C] Cognitive Modelling**	**[6D] Absolute Pitch and Language**	
Chair	Richard ASHLEY	Catherine STEVENS	Dawn BLACK	Cong JIANG	
08:30-9:00	Yuh-Wen WANG, Shan-Chuan TENG, Yunn-Wen LIEN, Yun-An HYANG, Jyh-Horng CHEN: Music Induces Mindfulness State: The Sujecheon Effect	Jonna VUOSKOSKI, Eric CLARKE, Charles SPENCE, Marc THOMPSON: Sight & Sound: Crossmodal Interactions in the Perception of Musical Performance	Ji Chul KIM, Edward W. LARGE: Pitch Dynamics on Multiple Time Scales in a Neurodynamic Model of Melodic Perception	Ao CHEN, Rene KAGER: Split between Musical Pitch and Speech Pitch Processing, Evidence from Early Infancy	
09:00-09:30	Jeong Mi PARK, Suk Won YI, Chun Kee CHUNG, June Sic KIM, Kyung Myun LEE, Jaeho SEOL: Musical Context Changes Neural Responses to Tonic Triad	Robert FULFORD, Carl HOPKINS, Gary SEIFFERT, Jane GINSBORG: Sight, Sound and Synchrony: Effects of Attenuating Auditory Information on Duo Violinists' Behaviours in Performance	Gerald GOLKA, Werner GOEBL: Tracking Expressive Performances with Linear and Non-Linear Timing Models	Kelly JAKUBOWSKI, Daniel MüLLENSIEFEN, Lauren STEWART: A Developmental Study of Latent Absolute Pitch in UK Children	
09:30-10:00	Niklas BÜDENBENDER, Gunter KREUTZ: Neuronal Processes during Identification of Familiar Melodies: Influences of Musical Expertise and Syntactic Structures	Morton WAN: Embodied Meanings in Beethoven's Late Piano Sonatas: A Performer Speaks	Martin HARTMANN, Petri TOIVIAINEN, Olivier LARTILLOT: Perception of Segment Boundaries in Musicians and Non-Musicians	Xiaonuo LI: Absolute Pitch among Students at the Shanghai Conservatory of Music: A Large-Scale Direct-Test Study	
10:00-10:30	Jung Nyo KIM, Ji Chul KIM, Edward LARGE: Oscillatory Brain Dynamics in the Processing of Implied Harmony	Randall HARLOW: Psychohaptics and the Affordances of Musical Instruments: Toward an Ecological Theory of Performance	Satoshi MORIMOTO, Gerard B. REMIJN, Yoshitaka NAKAJIMA: Computational Model-Based Analysis of Context Effects on Chord Processing	Joey L. WEIDEMA, M. Paula RONCAGLIA-DENISSEN, Henkjan HONING: The Effect of Musical Pitch on the Perception of Lexical Pitch in Speakers of Mandarin	
10:30-11:00	**Break**				
11:00-12:30	**Keynote 3 & Young Researcher Award 1 – SEOUL** John RINK Chair Keynote 3 Eric CLARKE: Music, Empathy and Cultural Value Young Researcher Award 1 Lauren HADLEY: The Effects of Anomaly on Music Reading: Evidence from Eye Movements				
12:30-14:00	**LUNCH**				

Room	SEOUL	KYOTO	MONTREAL	LIÈGE	LOS ANGELES
Session 7	**[7A] Symposium 2**	**[7B] Cognitive Neuroscience of Music 3**	**[7C] Affections in Musical Performance**	**[7D] Perception of Pitch**	**[7E] Schemata**
Chair	Rebecca SCHAEFER	Christoph REUTER	Gunter KREUTZ	Edward W. LARGE	Robert GJERDINGEN
14:00-14:30	***Musical Movement: Effects of Cueing and Feedback*** Discussant: Peter KELLER Rebecca SCHAEFER, Alexa M. MORCOM, Neil ROBERTS, Katie OVERY: fMRI Measures of Cued Movement Nicolas FARRUGIA, Simone DALLA BELLA: Tempo Effects on Movement Kinematics: A Marker of Expertise in Professional Drummers Marc THOMPSON, Tommi HIMBERG, Justin LONDON, Petri TOIVIAINEN: Effect of Tempo and Vision on Interpersonal Coordination of Timing Floris Van VUGT, Eckart ALTENMUELLER, Barbara TILLMANN: The Brain Uses Auditory Feedback for Motor Learning of Timing in Healthy and Stroke Populations	Tom COLLINS, Daniel A. ABRAMS, Rohan CHANDRA, Christina YOUNG, Andreas ARZT, Vinod MENON: Neural Tracking of Musical Motives Revealed by a Combination of FMRI and Music Information Retrieval Techniques	Emery SCHUBERT, Dorottya FABIAN, Renee TIMMERS: Defining Expressiveness in Music: Cognitive and Cross-Cultural Perspectives	Karl LERUD, Ji Chul KIM, Edward W. LARGE: A Neurodynamic Account of Residue Pitch	Claire ARTHUR: Does Harmony Affect Scale Degree Qualia? A Corpus Study Investigating the Relationship of Scale Degree and Harmonic Support
14:30-15:00		Frank RUSSO: Understanding the Perception of Music from the Perspective of Oscillatory Activity within and between Brain Networks	Hauke EGERMANN, Alexander FOERSTEL, Stephen MCADAMS:: The Effects of Continuous Subjective Music Experience Ratings on Physiological and Psychological Measures of Activation and Valence	Song Hui CHON, David HURON: Does Auditory Masking Explain the High Voice Superiority?	So-Young LEE, Elizabeth Park: Humane Tempo vs. Metronomic Tempo
15:00-15:30		Blair KANESHIRO, Jacek DMOCHOWSKI, Anthony NORCIA, Jonathan BERGER: Toward an Objective Measure of Listener Engagement with Natural Music Using Inter-Subject EEG Correlation	Petri TOIVIAINEN, Elvira BRATTICO: Neural Substrates of Perceived Musical Emotion during Continuous Listening	Xuejing LU, William Forde THOMPSON: Spatial Representations of Pitch among Individuals with Congenital Amusia	
15:30-16:00		Kirsteen DAVIDSON-KELLY, S. HONG, R.s. SCHAEFER, E. J. VAN BEEK, N. ROBERTS, K. OVERY: Expert Musical Imagery: An FMRI Investigation	Mary BROUGHTON, Emery SCHUBERT, Catherine STEVENS, Dominic HARVEY: Continuous Self-Report Engagement Responses to the Live Performance of a Post-Serialist Solo Marimba Work	Diana DEUTSCH, Cong JIANG, Trevor HENTHORN, Shibin ZHOU : The Tritone Paradox with Chinese Children Aged 12 and 13	
16:00-16:15	**Break**				

	Poster Session 3 - Baekyang Building Lobby		
16:15-17:00	[3-01] William RANDALL, Nikki RICKARD: Emotional Reasons for Personal Music Listening: Frequencies and Affective Outcomes [3-02] Finn UPHAM: Listener Respiration: Active, Attentive, or Distracted [3-03] Mio SAKURAI, Masanobu MIURA: Automatic Estimation of Released Decades of Japanese-Anime Music from Musical Audio [3-04] Lira YU, Masaki TOMONAGA: An Experimental Study about Behavioral Synchrony between Chimpanzees [3-05] Takeharu SENO: Music Modulates the Strength of Illusory Self-Motion Perception (Vection) [3-06] Steven MORRISON, Harry PRICE, Eric SMEDLEY, Cory MEALS: Gestural Content Influences Evaluations of Ensemble Performance [3-07] Birgitta BURGER, Suvi SAARIKALLIO, Marc THOMPSON, Geoff LUCK, Petri TOIVIAINEN: That's the Way I like It: Relationships Between Musical Genre And Dance Preferences and Music-Induced Movement [3-08] Hiroshi KAWAKAMI, Yuki MITO, Toru OZAWA, Yukitaka SHINODA, Toshihiro IRIE, Mieko MARUMO: Tempo Analysys of Dance Movements: The Master and a Disciple of Nihon Buyo [3-09] Neta COHEN-SHANI, Zohar EITAN, Dominique LAMI: Between Peaks and Troughs: Local and Global Aspects of Contour and the Perception of Melodic Direction	[3-10] Claudia LAPPE, Markus LAPPE, Christo PANTEV: Differences in Brain Responses to Melodic and Rhythmic Deviations in Musicians [3-11] Iballa BURUNAT, Petri TOIVIAINEN, Vinoo ALLURI, Brigitte BOGERT, Tapani RISTANIEMI, Mikko SAMS, Asoke NANDI, Elvira BRATTICO: Musical Feature Processing In The Brain During A Naturalistic, Continuous Listening: A Replication Study [3-12] Dahye BAE, June Sic KIM, Kyung Myun LEE, Chun Kee CHUNG, Suk Won YI: Musical Training Reflected in Beta-Band Activity to Meter [3-13] Xiaoluan LIU: The Impact of Emotions and Finger Stretch on the Dynamics of Piano Music [3-14] Carmen Fregoneze: Musical and Technical Demands in Piano Works by Two Brazilian Composers [3-15] Satoshi MIYAWAKI, Hisataka KATOU, Masanobu MIURA: Proficiency Estimation for Piano Performance of Burgmueller's "la Candeur" [3-16] Z. Funda YAZICI, Esra MUNGAN, Mustafa Uğur KAYA: Gestalt Universals or Musical Enculturation: Perceiving Boundaries in Unfamiliar Turkish Makam Music [3-17] Nancy GARNIEZ: Can Individuals, Both Trained and Untrained, Be Made Aware of Their Involuntary Responses to Tone Relatedness and, If So, Would It Make any Difference?	[3-18] Chie OHSAWA, Satoshi OBATA, Takeshi HIRANO, Minoru TSUZAKI, Hiroshi KINOSHITA: Flexible Modulation in Spatial Representation of Piano Keyboard in Trained Pianists [3-19] Shengchen LI, Dawn BLACK, Elaine CHEW, Mark D. PLUMBLEY: Evidence that Phrase-Level Tempo Variation May be Represented Using a Limited Dictionary [3-20] Christopher BARTLETTE, Anson JABLINSKI: Identifying Structure in a (16,9) Hyperdiatonic System [3-21] Shuo CHEN: Analysis on the Method and Effectiveness of Playing the Piano of a Child with Autism [3-22] Scott MURPHY: Using Discrete Calculus to Correlate a Sinusoidal Model of Meter with Normative Rhythms and Structural Accents [3-23] Hyun Ju CHONG, Sung-Rae CHO, Soo Ji KIM: Hand Rehabilitation Using Midi Keyboard Playing in Adolescents with Brain Damage: A Preliminary Study [3-24] Justin LONDON, Nori JACOBY, Rainer POLAK: The Effect of Ensemble Size, Musical Role, and Individual Performer on Collective Entrainment in Khasonka Drumming From Mali [3-25] Hauke EGERMANN, Cornelius STIEGLER: Communicating National and Temporal Origin of Music: An Experimental Approach to Applied Musical Semantics

Room	SEOUL	KYOTO	MONTREAL	LIÈGE	LOS ANGELES
Session 8	**[8A] Musical Practice**	**[8B] Music and Memory 2**	**[8C] Cross-Cultural Studies of Music**	**[8D] Singing 1**	
Chair	Jane GINSBORG	Ji Chul KIM	Kyung Il KIM	Rytis AMBRAZEVIČIUS	
17:00-17:30	Friedrich PLATZ, Reinhard KOPIEZ, Andreas C. LEHMANN, Anna WOLF: When Practice Makes Perfect: The Influence of Deliberate Practice on Skill Acquisition in Musicians: A Meta-Analysis	Richard ASHLEY: Long Term Memory for Musical Form: A Naturalistic Investigation Using Popular Songs	Patrick SAVAGE, Steven BROWN, Emi SAKAI, Thomas CURRIE: Statistical Universals Reveal the Structure and Functions of Human Music	Priyanka SHEKAR, Takako FUJIOKA: The Effects of Timbre and Musical Training on Vocal Pitch Matching Accuracy	
17:30-18:00	Towaki SHOSHI, Yoko OGAWA: An Analysis of Self-Evaluation Reports of the Practice Process of University Students Who Are Piano Beginners	Reinhard KOPIEZ, Friedrich PLATZ, Anna WOLF: Is There a Music-Stimulated Enhancement Effect on Autobiographical Memory? Towards Schulkind et al.'s (1999) Hypothesis of a Reminiscence Bump	Daniel SHANAHAN, Eva SHANAHAN: The Densmore Collection of Native American Songs: A New Corpus for Studies of Effects of Geography, Language and Social Function on Folksong	Mayumi ADACHI, Sanae OGURA: Exploration of Singing Skills in Preschool Children	
18:00-18:30	Rosane CARDOSO DE ARAUJO, Margaret AMARAL DE ANDRADE: A Study about Flow Theory and Teenager's Musical Practice	Michael W. WEISS, E. Glenn SCHELLENBERG, Sandra E. TREHUB, Emily J. DAWBER: Processing Advantage for Vocal Melodies: Are Children Like Adults?	Maria WITEK, Jingyi LIU, John KUUBETERZIE, Senyo ADZEI, Appiah Poku YANKERA, Morten KRINGELBACH, Peter VUUST: Cognitive, Affective and Embodied Responses to Rhythmic Complexity: A Cross-Cultural Comparison of West African and North American Non-Musicians	Steven DEMOREST, Peter PFORDRESHER, Jamey KELLY: School Music Participation: Exploring the Role of Students' Self-Concept and Singing Ability	
18:30-20:00	**BANQUET – Grand Hilton**				

Friday 8 August					
Room	**SEOUL**	**KYOTO**	**MONTREAL**	**LIÈGE**	**LOS ANGELES**
Session 9	**[9A] Workshops**	**[9B] Music Therapy: Rhythm and Motor**	**[9C] Musical Ability**	**[9D] Ears and Eyes**	
Chair		Soo Ji KIM	Yeajin KIM	Johanna DEVANEY	
08:30-09:00		Konstantinos TROCHIDIS, Hamzeen HAMEEM, Simon LUI: Motor Rehabilitation for Stroke Patients Using Musical Rhythmic Stimulation			
09:00-09:30	Jun Young KWAK: Awakening the Kinesthetic Senses in Piano Performance	Päivi JÄRVINEN-LEPISTÖ, Birgitta BURGER, Esa ALA-RUONA: Motor Performance in Post-Stroke Recovery Using Active Music Therapy	Yuko ODAHARA, Yumi HOSHINO: Musical Ability to Search for Essential Meaning in Appreciation Task: The Case of Minimal Music	Teruo YAMASAKI: The Effect of Exposure Time on the Intermodal Effect of Audio-Visual Stimulus: Comparison between Effect of Music and Effect of Paintings	
09:30-10:00	Megha MAKAM, Blair KANESHIRO, Jonathan BERGER: An Interactive Introduction to the Music and Movements of Brazilian Capoeira	Florina SPETH, Uwe SEIFERT: Robot-Assisted Arm Training and Rhythmic Acoustic Stimulation: The Role of Different Rhythmic Stimuli on Motivation and Performance in Motor-Rehabilitation Training After Stroke	Helen FARRELL: Redefining Music Excellence: Musical Futures in Special and Inclusive Education	Saki FUJIYAMA, Fumiyuki TAKISHITA, Toru YAHAGI, Shin-Ichiro IWAMIYA: Comparison of the Process for Creating Perceived Congruence between a Moving Picture and Sound Based on Synchronization of Auditory and Visual Accents and Similarity of Auditory and Visual Affective Impressions	
10:00-10:30		Loreen POGRZEBA, Margareta ERICSSON, Karina LARSSON, Markus WACKER, Bernhard JUNG: Towards a Vocabulary for Describing 3D Motion Data in Functionally Oriented Music Therapy	Anna WOLF, Reinhard KOPIEZ: When They Start Thinking in Sound: Assessing the Ear Training Skills of First-Year Music Education Students	Olivia M. PODOLAK, Mark A. SCHMUCKLER : The Influence of Tonality During Sight-Reading Performance	
10:30-11:00	**Break**				

Room	SEOUL	KYOTO	MONTREAL	LIÈGE	LOS ANGELES
Session 10	**[10A] Music and Memory 3**	**[10B] Preference and Analysis of Sound**	**[10C] Expectation, Education and Evaluation in Performance**	**[10D] Singing 2**	
Chair	Daniel SHANAHAN	Michael SCHUTZ	John RINK	Mayumi ADACHI	
11:00-11:30	John Ashley BURGOYNE, Jan VAN BALEN, Dimitrios BOUNTOURIDIS, Themistoklis KARAVELLAS, Frans WIERING, Remco C. VELTKAMP, Henkjan HONING: The Contours of Catchiness, or Where to Look for a Hook	Daisuke ISHIKAWA, Akira NISHIMURA, Nobuo KOIZUMI: Predicting Music Listening Preference from Impressions of Music Pieces under Four Moods	Kristin KUETER, Jonathan BERGER: Cherishing the Error: An Interplay between Musical Schemas and the Economic Heuristics that Guide Our Everyday Behavior	Dawn BLACK: Automatic Identification of Emotion in Chinese Opera Singing	
11:30-12:00	David SEARS , William E. CAPLIN, Stephen MCADAMS: Expecting the Classical Cadence: Priming Effects	Yoshitaka NAKAJIMA, Takayuki SASAKI, Kazuo UEDA, Gerard B. REMIJN: Auditory Grammar in Music	Dirk MOELANTS, Luc NIJS: Evolution of Timing through Instrumental Practice by Violinists and Clarinetists	Rytis AMBRAZEVIČIUS: Singing Indoors and Singing Outdoors: Formant Tuning	
12:00-12:30	Kat AGRES, Marcus PEARCE : An Information-Theoretic Approach to Learning, Memory, and Expectation in Unfamiliar Musical Styles	Nikki MORAN, Lauren HADLEY, Peter KELLER: Observers Can Recognise Real Versus Fake Free-Improvisation Duos through Nonverbal Backchannel Cues	Florian ECKL, Erica BISESI, Anders FRIBERG, Richard PARNCUTT: A Computational Model of Immanent Accent Salience in Tonal Music: Perceptual Evaluation	Philip FINE, Jane GINSBORG, Christopher BARLOW: Making Myself Understood: Factors Affecting the Understanding of Sung Text	
12:30-14:00	**LUNCH**				

Room	SEOUL	KYOTO	MONTREAL	LIÈGE	LOS ANGELES
Session 11	**[11A] Sadness in Music**	**[11B] Music in Digital Age**	**[11C] Sound beyond Media**	**[11D] Music and Language**	**[11E] Listening and Attention**
Chair	David HURON	Olivier LARTILLOT	Yoshitaka NAKAJIMA	Hye-Yoon CHUNG	Michael OEHLER
14:00-14:30	Henna-Riikka PELTOLA, Jonna VUOSKOSKI, Tuomas EEROLA: Why Listen to Sad Music? Exploring the Underlying Reasons	Ichiro FUJINAGA, David SEARS, Andrew HANKINSON: Big Data for the Music Perception and Cognition Community	Victoria WILLIAMSON, Michael SOUTH, Daniel MÜLLENSIEFEN: Sound Quality Enhances the Music Listening Experience	Ming Wai PANG: Tonal and Segmental Perception in Native Cantonese-Speaking Musicians, Amateur Musicians and Non-Musicians	Song Hui CHON, David HURON: Instrument Identification in Concurrent Unison Dyads: The Effect of Timbre Saliency
14:30-15:00	Evan COPELAND, Kirsten NISULA, David HURON: Sad Lyrics are Sung Slower: The Impact of Arousal, Not Valence, on Syllable Rate	Marco SCIREA, Yun-Gyung CHEONG, Byung-Chull BAE: Mood Expression in Real-Time Computer Generated Music Using Pure Data	Meghan GOODCHILD, Dominique BEAUREGARD CAZABON, Jamie WEBBER, Jonathan WILD, Stephen MCADAMS: (Re)Orchestrating Timbre: Behavioural and Psychophysiological Responses to Changes in Orchestration	Suki S. Y. YIU: Linguistic and Musical Scaling of Cantonese Tones	Jiaxi LIU: Perceptual Salience of Cadential and Non-Cadential Glides in Semitone Sequences
15:00:15:30		Adam STRANGE, Atsushi MARUI: An Investigation of the Cross Modal Interaction between Songs and Their Album Covers in the Mobile Digital Environment	Chelsea DOUGLAS, Jason NOBLE, Stephen MCADAMS: Soundmass and Auditory Streaming: A Perceptual Study of Ligeti's Continuum	M. Paula RONCAGLIA-DENISSEN, Fleur BOUWER, Henkjan HONING: Investigating the Integration of Long-Distance Dependencies in Language and Music	Sarah SAUVE: Effects of Musical Training on Auditory Grouping
15:30-16:00		Sungkyun CHANG, Kyogu LEE: Saliency-Driven Model for Perceptual Audio Onset Detection	Yuriko Hase KOJIMA: Finding Meanings between Instrumental Sound and Taped Sound: Listening Analysis of "Water Music" by Toru Takemitsu	Son-Hwa CHANG: Constitution of Musical Meaning: A Pragmatic Approach to Constitutions of Musical Sign and Its Utterance Meaning on Basis of Intramusical Context	Mariko HAMAMURA, Shin-Ichiro IWAMIYA: Difference of Optimum Listening Level of Music and Perceived Loudness of Sound between Men and Women
16:00-16:30	**Break**				
16:30-18:00	**Keynote 4, Young Researcher Award 2 & Closing Ceremony – SEOUL** Ian QUINN, Chair Keynote 4 David HURON: A Theory of the Musically Sublime Young Researcher Award 2: Sook Young WON, Simulation of One's Own Voice in a Two-Parameter Model				

ICMPC 13-APSCOM 5
Proceedings

Edited by Moo Kyoung Song

Copyright © 2014 ICMPC 13 – APSCOM 5
Copyright of the content of an individual abstract is held by the first named (primary) author of that abstract.
All rights reserved.

Table of Contents

Table of Contents 2

Monday 4 August 2014

Welcome / Keynote Address 1 4
[1A] Music and Emotion 4
[1B] Music in Everyday Life 10
[1C] Music and Movement 1 11
[1E] Acoustics and Psychoacoustics 20
[2A] Music and Timbre 21
[2B] Listening as a Creative Process 27
[2C] Musicking 41
[2D] Music and Movement 2 46
[2E] When Music Delivers Pleasure 65

Tuesday 5 August 2014

[3A] Cognitive Neuroscience of Music 1 69
[3B] Rhythm and Timing 70
[3C] Musical Structure 76
[3D] Social Psychology of Music 86
[3E] Power of Music 96
Keynote Address 2 105
[4A] Symposium 1 105
[4B] Music and Movement 3 109
[4C] Music and Health 112
[4D] Music Performance, Expression and Objective Analysis 113
[4E] Music and Emotions 127
[5A] Tonality vs Atonality 132
[5B] Musical Schemata and Cognition 145
[5C] Rhythm in Language, Music and Movement 146
[5D] Music and Memory 1 148

Thursday 7 August 2014

[6A] Cognitive Neuroscience of Music 2 154
[6B] Multimodal Perception and Performance 158
[6C] Cognitive Modeling 164
[6D] Absolute Pitch and Language 171
Keynote Address 3 & Young Researcher Award 1 172
[7A] Symposium 2 175
[7B] Cognitive Neuroscience of Music 3 177
[7C] Affections in Musical Performance 183

[7D] Perception of Pitch 185
[7E] Schemata 194
[8A] Musical Practice 200
[8B] Music and Memory 2 204
[8C] Cross-Cultural Studies of Music 205
[8D] Singing 1 209

Friday 8 August 2014

[9A] Workshops 214
[9B] Music Therapy: Rhythm and Motor 214
[9C] Musical Ability 228
[9D] Ears and Eyes 232
[10A] Music and Memory 3 238
[10B] Preference and Analysis of Sound 239
[10C] Expectation, Education and Evaluation in Performance 245
[10D] Singing 2 250
[11A] Sadness in Music 259
[11B] Music in Digital Age 263
[11C] Sound beyond Media 271
[11D] Music and Language 277
[11E] Listening and Attention 289
Keynote Address 4 & Young Researcher Award 2 301

Poster Session

Poster Session 1 311
Poster Session 2 344
Poster Session 3 387

Author Index 417

Welcome / Keynote Address 1

SEOUL 09:00 – 10:30 Monday 4 Aug 2014

ESSE EST PERCIPI? SOME EARLY MODERN PERSPECTIVES ON THE FUNDAMENTAL MIND-DEPENDENCY OF OBJECTS

Sukjae Lee
Seoul National University
Opening, Time: 09:00

The ontological status of the objects of our cognition has been of perennial interest to philosophers of the Western philosophical tradition. While some endorsed the commonsensical view that objects exist in their own right, regardless of whether they are perceived, others emphasized the mind-dependent nature of objects, insisting that their very being is fundamentally grounded in the fact that some mind perceives them. Roughly two hundred fifty years ago, the latter approach emerged in a prominent manner in that a number of key philosophers held that perception is integral to what it is to be an object. That is, in the late seventeenth and eighteenth centuries, this idea that objects of our cognition are fundamentally dependent on the mind of the perceiver emerged as a mainstay of the philosophical terrain. While the prominent role of the mind and its perception might seem foreign or counterintuitive to our contemporary sensibilities, such central figures as George Berkeley, Gottfried Wilhelm Leibniz, and Immanuel Kant all vigorously propounded this view. We, thus, are naturally led to wonder what reasons prompted them to espouse this fundamental role of perception. As expected, however, while these philosophers all accepted *that* objects are fundamentally mind-dependent, they differed in how each of them articulated the fundamental *nature of this dependence*. That is, how objects depend on the mind turn out to be rather different and distinctive for each philosopher. This paper aims to examine the common and distinctive features of the views of these philosophers, and consider what implications their views might have for how we regard the objects of cognition. A central question guiding our examination will be in what sense the three philosophers aforementioned took the objects of our cognition to be 'phenomena'—as "things that appear to us." The notion of 'phenomenon' will itself require clarification, since the very description of "a thing that appears" raises further questions. Does it imply that there is a thing behind, as it were, the appearance, which emerges as it appears? Or are appearances all there are, the only things that we perceive? In pursuing such questions, we will set into relief the distinctive positions Berkeley, Leibniz, and Kant occupy as they promote the view that the ordinary objects of our cognition are inseparably dependent on the fact that we perceive them.

Session 1

[1A] Music and Emotion

SEOUL 14:00-16:00 Monday 4 Aug 2014

RESEARCHING THE ORIGINS OF MUSIC WITH METHODS FROM FIRST LANGUAGE ACQUISITION RESEARCH

Kathryn Davies
Hong Kong Academy of Performing Arts, Hong Kong, China
Time: 14:00

With the growing interest in the origins of music (e.g. Brown et al., 2000; Patel, 2008; Mithen, 2009), there has been much discussion on useful approaches for researching the role music plays with human evolution (e.g. Rebuschat, Rohrmeier, Hawkins, Cross, 2012). However, within the field of language and developmental psychology, investigations into language evolution are well underway. A focal point of this research is first language acquisition. Recently, researchers have developed new approaches to study language acquisition for which cultural cognition plays a key role (Tomasello, 1999; Tomasello et al., 2005). According to Tomasello (2003) there are three additional learning mechanisms that play an important part of language acquisition: entrenchment and competition, schematization, and distributional learning (i.e. learning through the statistical properties inherent in language). Thus far, there is emerging evidence that music production or perception relies on cultural cognition (e.g. joint attention and shared intentionality) as well as distributional learning. In this paper, I argue that as methodology for empirical research on the origins of music and music acquisition develop, it is valuable to rely on an established framework from linguistics on ontological language development. Moreover, Tomasello's (2003) framework offers a lot of potential for music research. Despite differences in how music and language function and are used, bringing together these two areas of research can offer valuable insights on the ontological and phylogenic development of music and language.

PITCH CENTRICITY AS AN EVOLUTIONARY INNOVATION OF VOCAL COMMUNICATION

Piotr Podlipniak
Department of Musicology, Adam Mickiewicz University, Poznan, Poland
Time: 14:30

ABSTRACT

Pitch centricity belongs to the most ubiquitous features of music. The process of establishing a pitch center is possible thanks to an implicit analysis of the pitch occurrence frequency in sound sequence. This analysis enables the listeners to create more or less successful expectations concerning the appearance of a pitch center. The appearance of the pitch center is additionally accompanied by the feeling of resolution or relaxation whereas the appearance of distant pitches in tonal hierarchy triggers tensions or anxiety. As far as we know,

there is nothing similar to pitch centricity in other domains of human cognition as well as in animals' vocal communication. Thus, pitch centricity seems to be the unique and species-specific feature of human vocal communication, which suggests its evolutionary origin. Scientists have usually explained an intuitive character and ubiquity of pitch centricity as a result of general principles of cognition. The aim of the paper is to present another point of view in which pitch centricity is claimed to be the adaptive innovation of vocal communication which emerged in the biological evolution of hominines and which is based on an evolutionarily new mental mechanism. It is proposed in the presentation that in the course of hominine evolution the ability of pitch center recognition became an adaptive innovation which enabled a more effective social consolidation by means of hierarchically organized pitch sequences. It is also suggested that the origin of this ability has its roots in the 'Baldwin effect'. It means that at the beginning pitch centricity was a cultural invention proliferated by means of social learning. However, an instinctive recognition of pitch center which emerged accidentally was preferred by the cultural environment as an important source of evolutionary selection. This process led to the proliferation of a predisposition to join three previously separate abilities–the implicit recognition of the frequency of pitch occurrence, working memory and the emotional assessment of predicted stimuli–into a new mental capability which enables people to intuitively recognize a pitch center. As a result pitch centricity became an innovative tool of vocal communication.

1. INTRODUCTION

Human vocal communication is rich in the variety of tools such as crying, laughter, speech, and singing. From all these forms of communication, speech as a vocal expression of language and singing as a vocal form of music belong to the most complex acoustic phenomena. They seem also specific solely to humans, which indicates that some of their components should be the evolutionary innovations (Fitch and Zuberbühler, 2013). However, because of evolutionary opportunism, music and language have to contain non-innovative, evolutionarily old elements too. Indeed, they are composed both of affective prosody whose some features are shared with other mammalian taxa (Zimmermann, Leliveld, and Schehka, 2013), and discrete elements which form the generative codes called 'Humbolt systems' (Merker, 2002).

Since both music and speech possess an affective component but only language contains the propositional semantics, Darwin (1871) suggested that the pre-semantic language, which–in his opinion–preceded the evolution of contemporary language, had to be musical in character. Because the segmental phonology of speech differs substantially from the segmental phonology of music it is more probable that the 'musilanguage' (Brown, 2000) or the 'prosodic protolanguage' (Fitch, 2006; 2013), from which contemporary language and music evolved, was composed of continuously sung phrases rather than discrete pitch classes as it is the case in contemporary music. The difference between the phonological systems of speech and music consists in the distinctive features of phonemes and their musical counterparts (Patel, 2008). While in the case of speech phonemes their main distinctive features are formats (Berent, 2013) in music the crucial role in discrimination of the discrete structural components plays pitch and the temporal arrangement of sounds based on a steady beat. Although in some languages pitch differences (in discrete level tone languages) and temporal proportions between phonemes (in quantity sensitive languages) influence the comprehension of speech, only in music these features become the main basis of syntactic rules. One of the ubiquitous elements of music syntax which is absent from speech and which is related to pitch classes arrangement is pitch centricity (Huovinen, 2002). Pitch centricity is also indicated as an example of musical universals (Brown and Jordania, 2013).

2. PITCH CENTRICITY AS A UNIQUE CHARACTERISTIC OF HUMAN VOCAL COMMUNICATION

The combination of pitches is present not only in human vocal communication. The most obvious examples of a vocal communication composed of these elements are songbirds' songs which resemble in this regard human music (Rothenberg et al., 2014). Songbirds similarly to humans are vocal learners and they learn their songs' dialects most effectively during the sensitive periods (Moorman and Bolhuis, 2013). Thus, songbirds' songs are culturally transmitted combinations of rhythms and pitches in the same way as music is (Rothenberg et al., 2014). Moreover, unlike language and similarly to music they do not have propositional semantics (Marler, 2000). Also the affective component of songs processing in birds' brains seems to engage the same mesolimbic reward system as in the human brains during listening to music (Earp and Maney, 2012; Rothenberg et al., 2014). Some birds are even able to learn simple syntax (Gentner et al., 2006; Okanoya, 2013). However, pitch syntax in tonal music is much more complex than in every known songbird's song and vocal communication systems of other animals (Fitch and Zuberbühler, 2013). The organization of pitches in tonal music is based on a set of pitch classes–well recognizable for people raised in particular musical culture–combinations of which form pitch intervals. As far as it is known, there is no consistent use of harmonic intervals among birds' songs (Araya-Salas, 2012). What is more important, pitch syntax in music is hierarchical (Lerdahl and Jackendoff, 1983) and pitch centricity seems to be the crucial element necessary to create this hierarchy. There is nothing similar to pitch centricity in any known vocal communication systems observed in the animal kingdom as well as in other form of human vocal expressions. Therefore, if the rules of pitch center establishing are a result of inborn predisposition shared by all humans pitch centricity should be understood as an evolutionary innovation of vocal communication specific to *Homo sapiens*.

Pitch center plays the most prominent position in every tonal hierarchy which becomes a scaffold for the comprehension of musical pitch structure. The hierarchy is built by means of statistical learning during a lifelong exposition to music (Krumhansl, 1990) and serves as an implicit knowledge stored in a long-term memory (Krumhansl, 2004). Statistical learning is a ubiquitous form of implicit learning observed in many domains of human cognition (Kirkham, Slemmer, Johnson, 2002) as well as among other species (Conway and Christiansen, 2001; Hauser, Newport, and Aslin, 2001; Newport, et al., 2004). However, the application of the statistical learning in the creation of pitch hierarchy does not explain the reasons behind this mental hierarchy (cf. Conway and Christiansen, 2001) which is specific solely to music. In other words, the statistical learning alone seems not to be sufficient so as to elicit an experience of pitch centricity. After all, the phonemes of particular language occur with different frequencies during speaking similarly to pitch classes in

music. But there is nothing similar to pitch centricity in speech which one can name phonemic centricity. Another popular explanation of the specific character of tonal hierarchy is related to the general mechanism of prediction and emotional reaction to predicted stimuli.

The implicit knowledge of pitch classes' statistics allows listeners the creation of predictions concerning the successive pitch classes' occurrence. As a result, different probabilities of the pitch classes' occurrence in other pitch class contexts cause various emotional reactions in response to fulfilling or not fulfilling the predictions (Meyer, 1956; Huron, 2006). According to Huron (2006), the differentiation of subtle emotional reactions depends on the different probabilities. Thus, the emotions which are actually a result of prediction are misattributed to particular scale degrees. These subtle emotional characteristics are often treated as scale degrees' *qualia* (Huron, 2006; Margulis, 2014). However, also this explanation does not elucidate the specificity of the phenomenon of pitch centricity. Since the prediction is a general mechanism of cognition (Llinás, 2002), why the emotions caused by predictions are not misattributed to phonemes or other sound categories? Moreover, the emotional component of pitch center experience seems to be a very important factor which plays crucial role only in music syntax. These observations suggest that pitch centricity is a result of a domain specific mental tool rather than a byproduct of general cognition. If it is true pitch centricity had to fulfil an adaptive function in human origin.

3. PITCH CENTRICITY, VOCAL IMITATION, AND SPECTRAL SYNCHRONY

Singing is observed in all known human cultures in the world (Nettl, 2000). As a form of music which does not necessitate any technology (as in the case of instrumental music) singing was probably the most primitive form of music (Mithen, 2006; Morley, 2013). It is reasonable to suppose that the fully developed contemporary singing with pitch class syntax evolved from more primitive form of vocal communication which was based on affective calls (Fitch, 2013). Vocal communication, because of its ability to manipulate other individuals' behaviors became an important component of the human ancestors' culture. One of the parts of the proto-humans' cultural environment developing to the greatest extent was ritual culture (Merker, 2012) which relies on learning by imitation (Jablonka and Lamb, 2005). As was previously indicated, similarly to songbirds, but not to nonhuman primates (Janik and Slater, 2000), human possess the ability of vocal learning which enables people to imitate other individuals' vocal expressions. This ability was indispensable for the evolution of speech, the development of which depends critically on learning by imitation.

Also the proliferation of songs is equally dependent on imitation. However, unlike speaking, singing is usually a collective and synchronous activity (Cross, 2012). It suggests that the adaptive value of singing is related to social character of our species. Indeed, a group consolidation (Roederer, 1984; Storr, 1990; McNeill, 1995; Kirschner and Tomasello, 2010) and signalization of a group consolidation (Hagen and Bryant, 2003; Hagen and Hammerstein, 2009) are often indicated as the adaptive functions of music. The main feature of music which is believed to be responsible for these functions is rhythm (McNeill, 1995). However, the alignment of humans' brain states which leads to the experience of consolidation and in consequence fosters social cohesion can be realized not only thanks to rhythmic synchronization but also by spectral synchrony (Bharucha, Curtis, and Paroo, 2012). One of the spectral characteristics of harmonic sound is its fundamental frequency (F0) and its partials which the sensation of pitch depends on (Roederer, 2008). Thus, so as to serve as an effective tool of consolidation, synchronous singing cannot be chaotic also in terms of pitch. Additionally, in order to sing synchronously people must be able to predict pitch sequences which will be sung by others. This is what probably forced our ancestors to memorize pitch sequences. One of the features which facilitates remembering pitch sequences is definitely pitch syntax based on pitch centricity. Therefore, the consolidating function of singing can explain the adaptive value of pitch centricity. However, it does not explain how the ability to recognize pitch center evolved.

4. PITCH CLASS SYNTAX AND THE 'BALDWIN EFFECT'

The most difficult question in the evolutionary scenario of pitch centricity evolution is how the ability to recognize pitch center could have evolved from scratch. The research of living primates gives basis to claim that our pre-musical ancestors were endowed with the abilities of statistical learning, creating predictions of heard sounds, emotional assessment and working memory. It is also reasonable to assume that they possessed auditory learning capacity (Jarvis, 2013) which enables them to associate and recognize sounds. As the research on rhesus monkeys (*Macaca mulatta*) suggests (Wright et al., 2000) most probably our predecessors had had the ability of octave generalization before they became vocal learners. All these abilities are necessary but not sufficient to create and recognize pitch sequences. Only the appearance of vocal learning allowed transmission of ritualized pitch sequences (Merker, 2005) similar to songbirds' songs. However, even at this stage of evolution the possibility to use syntactically complex pitch sequences was restricted. Something more was needed to develop intuitively comprehensible pitch syntax. I think that the key factor indispensable for using pitch classes in a generative way has been the predisposition to recognize pitch center. Being endowed with all aforementioned abilities, hominines needed only one step to become skilled users of pitch sequences in a generative way. This step consists in coupling three previously separate abilities–the implicit recognition of the frequency of pitch occurrence, working memory and the emotional assessment of predicted stimuli–into a new mental capability. The implementation of the existing abilities in a new way for an animal to gain a new cognitive tool is consistent with evolutionary, opportunistic logic (Jacob, 1977). However, it does not explain how the ability to recognize pitch center was selected in the case where any pitch sequences were not organized around central pitch?

In order to resolve this dilemma it seems useful to assume that before the appearance of pitch centricity our ancestors had invented a vocal ritual composed of pitch sequence. Social invention is nothing exceptional among primates (Jablonka and Lamb, 2005) so it is possible that hominines who were vocal learners used their inventive power in the domain of vocal expression (cf. Dor and Jablonka, 2001). Because of the aforementioned potential of vocal calls for the alignment of brain states, it is very probable that the invented vocal ritual served as a means of consolidation. This assumption seems to be supported by the observations of living chimpanzees. Vocal activity called 'pant-hooting' is used by chimpanzees when a member of a community rejoins the rest of a group (Geissmann, 2000), which resembles to some extent the supposed consolidating function of our ancestral ritual. As a result of the ritual function, the better

synchronized in singing the participants were the more consolidated they appeared to be. In order to be more efficient at spectral synchronization they need to know the pitch sequence sung by the rest of the group. This task necessitated strenuous learning of the whole pitch sequence by heart.

Because of working memory restrictions, the pitch sequence was probably composed of many repeating pitches so as to facilitate the process of memorization. However, even in such a case, without the ability to pitch center recognition the task resembled the learning of atonal melodies by contemporary humans. The more elaborate the pitch sequence was the more strenuous the learning had to be. Therefore, the appropriate performance of complex pitch sequence could have been additionally a measure of time devoted to learning. Time-consuming learning of the ritual, in turn, indicated that one was more attached to a community. If the tradition of the ritual was very important for the community it was preserved through many generations. In such circumstances the ritual became a selective environment. As a consequence of genetic variability which characterizes every next generation, one individual gained accidentally the inherited predisposition to instinctively learn pitch sequences based on statistical learning facilitated by emotional assessment. In this moment the 'Baldwin effect' started to act, which means that social invention (pitch centricity) which was proliferated by means of social learning changed into instinctively learned trait (Baldwin, 1896). This was possible because the quicker learners were preferred by natural selection. In a long run perspective, even small advantage in the ability to consolidate with the group resulted in the proliferation of the ability to recognize pitch center among the whole population.

The innovative character of pitch centricity which fulfilled the consolidating function of pitch sequence explains the reason behind the evolution of music syntax. In the proposed scenario pitch centricity became a tool of emotion control used to enhance social bonding. This function is absent from language syntax which evolved because of other reasons most probably related to communication of referential meaning (Dor and Jablonka, 2001). Even though many mental mechanisms used in processing of music syntax are shared with language syntax processing (Koelsch, 2005; Patel, 2003), the specific emotional component of pitch center recognition suggests the functional specificity of pitch class structure. The instinctive recognition of pitch center allows creating vast number of possible tonal hierarchies dependent on cultural invention. From this perspective Western tonality is only one possibility. In this regard music syntaxes resemble language grammars. Insofar as language grammars are constrained by some semantic regularities (Dor, 2000) music syntaxes are restricted by pitch centricity.

Although the proposed evolutionary scenario has a speculative character its implications allow to empirical verification in the future research. One of them is the comparison of emotional reaction to predicted and non-predicted (syntactically correct and non-correct) sequences of speech and melodies. Additionally, the suggested evolution of pitch centricity by means of Baldwinian mechanism is consistent with the fact that, on the one hand, pitch syntaxes differ cross-culturally; but, on the other hand, pitch centricity is such a ubiquitous music trait. This observation shows that although people are endowed with some music-specific mental predispositions social invention is an important factor in the elaboration of music syntaxes.

5. REFERENCES

Araya-Salas, M. (2012). Is birdsong music? Evaluating harmonic intervals in songs of a Neotropical songbird. Animal Behaviour, 84(2), 309–313. doi:10.1016/j.anbehav.2012.04.038

Baldwin, J. M. (1896). A new factor in evolution. American Naturalist, (30 June, July), 441-451; 536-553.

Berent, I. (2013). The phonological mind. Cambridge, New York: Cambridge University Press.

Bharucha, J. J., Curtis, M. E., & Paroo, K. (2012). Musical Communication as Alignment of Brain States. In P. Rebuschat, M. Rohrmeier, J. A. Hawkins, & I. Cross (Eds.), Language and music as cognitive systems (pp. 139–155). Oxford: Oxford University Press.

Brown, S. (2000). The "Musilanguage" Model of Music Evolution. In S. Brown, B. Merker, & N. L. Wallin (Eds.), The origins of music. [consists of papers given at a Workshop on the Origins of Music held in Fiesole, Italy, May 1997, the first of a series called Florentine Workshops in Biomusicology] (pp. 271–300). Cambridge, Mass. [u.a.]: MIT Press.

Brown, S., & Jordania, J. (2013). Universals in the world's musics. Psychology of Music, 41(2), 229–248. doi:10.1177/0305735611425896

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. Trends in cognitive sciences, 5(12), 539–546.

Cross, I. (2012). Music as a Social and Cognitive Process. In P. Rebuschat, M. Rohrmeier, J. A. Hawkins, & I. Cross (Eds.), Language and music as cognitive systems (pp. 315–328). Oxford: Oxford University Press.

Darwin, C. R. (1871). The descent of man and selection in relation to sex. London: John Murray.

Dor, D. (2000). From the autonomy of syntax to the autonomy of linguistic semantics: Notes on the correspondence between the transparency problem and the relationship problem. Pragmatics & Cognition, (8/2), 325–356.

Dor, D., & Jablonka, E. (2001). How language changed the genes: toward an explicit account of the evolution of language. In J. Trabant & S. Ward (Eds.), Trends in linguistics. Studies and monographs: Vol. 133. New essays on the origin of language (pp. 147–173). Berlin: New York; Mouton de Gruyter.

Earp, S. E., & Maney, D. L. (2012). Birdsong: Is It Music to Their Ears? Frontiers in Evolutionary Neuroscience, 4. doi:10.3389/fnevo.2012.00014

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. Cognition, 100(1), 173–215. doi:10.1016/j.cognition.2005.11.009

Fitch, W. T. (2013). Musical Protolanguage: Darwin's Theory of Language Evolution Revisited. In J. J. Bolhuis & M. Everaert (Eds.), Birdsong, speech, and language. Exploring the evolution of mind and brain (pp. 489–503). Cambridge, Mass: MIT Press.

Fitch, W. T., & Jarvis, E. D. (2013). Birdsong and Other Animal Models for Human Speech, Song, and Vocal Learning. In M. A. Arbib (Ed.), Strüngmann Forum reports. Language, music, and the brain. A mysterious relationship (pp. 499–539). Cambridge, MA: The MIT Press.

Fitch, W. T., & Zuberbühler, K. (2013). Primate Precursors to Human Language: Beyond Discontinuity. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), Series in affective science. Evolution of emotional communication. From sounds in nonhuman mammals to speech and music in man (1st ed., pp. 26–48). Oxford: Oxford University Press.

Geissmann, T. (2000). Gibbon Songs and Human Music from an Evolutionary Perspective. In S. Brown, B. Merker, & N. L. Wallin (Eds.), The origins of music. [consists of papers given at a Workshop on the Origins of Music held in Fiesole, Italy, May 1997, the first of a series called Florentine Workshops in Biomusicology] (pp. 103–123). Cambridge, Mass. [u.a.]: MIT Press.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. Nature, 440(7088), 1204–1207. doi:10.1038/nature04675

Hagen, E. H., & Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. Musicae Scientiae, 13(2 Suppl), 291–320. doi:10.1177/1029864909013002131

Hagen, E. H., & Bryant, G. A. (2003). Music and Dance as a Coalition Sygnaling System. Human Nature, (14/1), 21–51.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. Cognition, 78(3), B53. doi:10.1016/S0010-0277(00)00132-3

Huovinen, E. (2002). Pitch-class constellations: Studies in the perception of tonal centricity. Acta musicologica Fennica: Vol. 23. Turku: Suomen Musiikkitieteellinen Seura.

Huron, D. B. (2006). Sweet anticipation: Music and the psychology of expectation (1st ed.). Cambridge, Mass, London: MIT.

Jablonka, E., & Lamb, M. J. (2005). Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life. Life and mind. Cambridge, Mass: MIT Press.

Jacob, F. (1977). Evolution and Tinkering. Science, (196/4295), 1161–1166.

Janik, V. M., & Slater, P. J. (2000). The different roles of social learning in vocal communication. Animal Behaviour, 60(1), 1–11. doi:10.1006/anbe.2000.1410

Jarvis, E. D. (2013). Evolution of Brain Pathways for Vocal Learning in Birds and Humans. In J. J. Bolhuis & M. Everaert (Eds.), Birdsong, speech, and language. Exploring the evolution of mind and brain (pp. 63–107). Cambridge, Mass: MIT Press.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition, 83(2), B35-42.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children☆☆☆. Evolution and Human Behavior, 31(5), 354–364. doi:10.1016/j.evolhumbehav.2010.04.004

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. Current Opinion in Neurobiology, 15(2), 207–212. doi:10.1016/j.conb.2005.03.005

Krumhansl, C. L. (1990). Cognitive foundations of musical pitch. Oxford psychology series: Vol. 17. New York: Oxford University Press.

Krumhansl, C. L. (2004). The Cognition of Tonality – as We Know it Today. Journal of New Music Research, (33/3), 253–268.

Lerdahl, F., & Jackendoff, R. (1983). A generative theory of tonal music. Cambridge, Mass: MIT Press.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling Tonal Tension. Music Perception: An Interdisciplinary Journal, 24(4), 329–366. doi:10.1525/mp.2007.24.4.329

Lieberman, P. (2000). Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought. Perspectives in cognitive neuroscience. Cambridge, Mass: Harvard University Press.

Llinás, R. R. (2002). I of the vortex: From neurons to self. Cambridge, Mass, London: MIT.

Margulis, E. H. (2014). On repeat: How music plays the mind. Oxford, New York: Oxford University Press.

Marler, P. (2000). Origins of Music and Speech: Insights from Animals. In S. Brown, B. Merker, & N. L. Wallin (Eds.), The origins of music. [consists of papers given at a Workshop on the Origins of Music held in Fiesole, Italy, May 1997, the first of a series called Florentine Workshops in Biomusicology] (pp. 31–48). Cambridge, Mass. [u.a.]: MIT Press.

McNeill, W. H. (1997). Keeping Together in Time: Dance and Drill in Human History (Reprint.). Cambridge: Harvard University Press.

Merker, B. (2002). Music: The Missing Humboldt system. Musicae Scientiae, (6/3), 3–21.

Merker, B. (2005). The Conformal Motive in Birdsong, Music, and Language: An Introduction. Annals of the New York Academy of Sciences, 1060(1), 17–28. doi:10.1196/annals.1360.003

Merker, B. (2012). The Vocal Learning Constellation: Imitation, Ritual Culture, Encephalization. In N. Bannan (Ed.), Music, language, and human evolution (1st ed., pp. 215–260). Oxford: Oxford University Press.

Merker, B. (2009). Ritual Foundations of Human Uniqueness. In S. Malloch & C. Trevarthen (Eds.), Communicative musicality. Exploring the basis of human companionship (pp. 45–59). Oxford, New York: Oxford University Press.

Meyer, L. B. (1956). Emotion and meaning in music. [Chicago]: University of Chicago Press.

Mithen, S. J. (2006). The singing neanderthals: The origins of music, language, mind, and body. Cambridge, Mass: Harvard University Press.

Moorman, S., & Bolhuis, J. J. (2013). Behavioral Similarities between Birdsong and Spoken Language. In J. J. Bolhuis & M. Everaert (Eds.), Birdsong, speech, and language. Exploring the evolution of mind and brain (pp. 111–123). Cambridge, Mass: MIT Press.

Morley, I. (2013). The prehistory of music: Human evolution, archaeology and the origins of musicality. Oyford: Oxford University Press.

Nettl, B. (2000). An Ethnomusicologist Contemplates Universals in Musical Sound and Musical Culture. In S. Brown, B. Merker, & N. L. Wallin (Eds.), The origins of music. [consists of papers given at a Workshop on the Origins of Music held in Fiesole, Italy, May 1997, the first of a series called Florentine Workshops in Biomusicology] (pp. 463–472). Cambridge, Mass. [u.a.]: MIT Press.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. Cognitive Psychology, 49(2), 85–117. doi:10.1016/j.cogpsych.2003.12.002

Okanoya, K. (2013). Finite-State Song Syntax in Bengalese Finches: Sensorimotor Evidence, Developmental Processes, and Formal Procedures for Syntax Extraction. In J. J. Bolhuis & M. Everaert (Eds.), Birdsong, speech, and language. Exploring the evolution of mind and brain (pp. 229–242). Cambridge, Mass: MIT Press.

Patel, A. D. (2003). Language, music, syntax and the brain. Nature Neuroscience, 6(7), 674–681. doi:10.1038/nn1082

Patel, A. D. (2008). Music, language, and the brain. Oxford, New York: Oxford University Press.

Roederer, J. G. (1984). The Search for a Survival Value of Music. Music Perception: An Interdisciplinary Journal, (1/3), 350–356.

Roederer, J. G. (2008). The physics and psychophysics of music: An introduction (4th ed.). New York: Springer.

Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., & Tchernichovski, O. (2014). Investigation of musicality in birdsong. Hearing Research, 308, 71–83. doi:10.1016/j.heares.2013.08.016

Storr, A. (1993). Music and the mind (1st ed.). New York: Ballantine.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. Journal of experimental psychology, 129(3), 291–307.

Zimmermann, E., Leliveld, L., & Schehka, S. (2013). Toward the Evolutionary Roots of Affective Prosody in Human Acoustic Communication: A Comparative Approach to Mammalian Voices. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), Series in affective science. Evolution of emotional communication. From sounds in nonhuman mammals to speech and music in man (1st ed., pp. 116–132). Oxford: Oxford University Press.

Zuberbühler, K., & Fitch, W. T. (2013). Primate precursors to human language: Beyond discontinuity. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), Series in affective science. Evolution of emotional communication. From sounds in nonhuman mammals to speech and music in man (1st ed., pp. 26–48). Oxford: Oxford University Press.

SOCIAL COMMUNICATIVE FUNCTIONS OF THE SINGERS' FORMANT IN CHORAL MUSIC

Peter Keller[1], Rasmus Koenig[2]
[1]The Marcs Institute, University of Western Sydney, Australia, [2]The Max Planck Institute for Human Cognitive and Brain Sciences, Germany
Time: 15:00

1. Background: Music is a human communicative art that may have deep biological roots. Long before humans evolved the capacity to sing and dance together in groups, other animals—such as crickets, frogs, and crabs—engaged in synchronous group behavior. These synchronous displays, produced by males in order to attract migrating females, are termed the "beacon effect". Some accounts of the origins of human musical behavior have postulated similar functions related to courtship.

2. Aim: The goal of the current study was to test whether the beacon effect occurs in human musical behavior. To do so, we conducted acoustic analyses on recordings of a male choir singing to audiences in which females were present or absent. We were interested in whether the presence of females would affect the degree of synchrony between voices, their intensity, and their spectral composition.

3. Method: The voices of members of the St Thomas Boys Choir in Germany, one of the world's premier choral ensembles, were recorded with head-worn microphones as they performed a short concert program three times: first with an all-male audience, then with female peers in the audience, and finally with an all-male audience again. The program consisted of two chorales by J. S. Bach. The choir comprised 4 sopranos (aged 12-13 years), 4 altos (aged 12-16), 4 tenors (aged 16-18), and 4 basses (aged 16-19). The female peers were aged 15-16years.

4. Results: Acoustic analyses of the recordings revealed that the basses increased the energy in a high frequency band (2500-3500 Hz) of voice's spectrum when females were in the audience. The sopranos, altos, and tenors showed no such effect. The presence of females did not affect the synchrony of voices or their overall intensity.

5. Conclusions: As increased energy in the 2500-3500 Hz frequency band, known as the "singers' formant", adds brilliance and carrying power to the voice, our finding is consistent with a beacon effect. This effect may constitute an attempt by the more sexually mature boys, who presumably have higher testosterone levels, to produce attractive vocal sounds to establish a privileged social communication channel with female listeners.

SPONTANEOUS AUDITORY-MOTOR ENTRAINMENT DURING SELF-PACED TAPPING IN CHIMPANZEES AND HUMANS

Yuko Hattori[1], Masaki Tomonaga[1], Tetsuro Matsuzawa[1]
[1]Primate Research Institute, Kyoto University, Japan
Time: 15:30

This study provides evidence that chimpanzees and humans show similar rhythmic entrainment when they hear simple auditory beat as distractor stimuli. While self-paced tapping two keys of an electric keyboard alternately, we played isochronous auditory beats as distractor stimuli. When these beats were close to their preferred tempo, tapping onset spontaneously showed a consistent relationship with auditory stimulus onset. Thus, our result shows auditory-motor entrainment in chimpanzees as well as humans and a strong

connection between auditory and motor-related areas in the brain was already deeply rooted in the common ancestor of chimpanzees and humans six million years ago. Our result therefore challenges the "Complex vocal learning hypothesis" that vocal learning has led to a tight coupling between auditory input and motor output, because chimpanzees do not have complex vocal learning. Flexible synchronized tapping to a wide range of auditory beats in humans was found only following instructions to pay attention to the stimuli. It remains unclear whether chimpanzees are capable of learning to align their tapping flexibly to auditory rhythms, but the absence of reports on rhythmic synchronized activity in the field may suggest that intentional synchronized movement with attention to external rhythms might have emerged during human evolution.

[1B] Music in Everyday Life

KYOTO 14:00-15:30 Monday 4 Aug 2014

THE EFFECT OF POSITIVE BACKGROUND MUSIC ON COMPLIANCE

Naomi Ziv[1]
[1]*College of Management - Academic Studies, Israel*

Time: 14:00

The effect of positive background music on compliance.

1. Background: Studies on the effect of music on persuasion in commercial messages have shown that background music, through its effect on mood, associations, and cognitive processing, may lead to the acceptance of messages. Similar effects have been found on empathy and pro-social behavior. However, the possible negative consequences of these effects have rarely been studied.

2. Aims: The aim of the present study was to examine whether these effects may occur when participants are asked to comply with a request which harms a third party.

3. Method: 120 participants took part in the study (24 males, 96 females, mean age = 23.03, SD = 2.52). Participants were randomly assigned to four groups: no background music, positive, unfamiliar instrumental music, a positive, familiar song in English, and a positive familiar song in Spanish. The study was conducted individually. Participants were told the study dealt with the effect of background music on cognitive processing. Participants were given a simple cognitive task for 2 minutes with or without background music. After completing the task, the researcher asked the participant to do them a favor: They were told there is another participant, who came especially to participate in the study. They were told that she needs to participate in order to complete the college requirement to take part in studies. However, the researcher "doesn't feel like" seeing her. The participant was asked to call her and say that the researcher is away. Participants' consent or refusal to comply was noted.

4. Results: Chi square tests on compliance rates were almost significant (Chi square = 7.55 p = .056). 12 participants (40%) in the control group agreed to comply with the request. In the positive instrumental music group 17 participants (55.7%) agreed, in the English song, 21 (70%) participants agreed, and 21 (70%) in the Spanish song agreed. No differences in mood were found between the groups. However, within the familiar English song group, participants who agreed to comply were less sad ($t = -2.12$, $p = .045$), and less stressed ($t = -3.51$, $p = .002$).

5. Conclusions: Results show that positive background music may influence people's behavior and lead them to comply more easily with requests, even those entailing harm to a third person.

CONTEXTUALIZED LISTENING: PLAYLISTS FOR EVERYDAY SITUATIONS

Amanda Krause[1]*, Adrian North*[1]
[1]*Curtin University, Australia*

Time: 14:30

1. Background: Everyday music listening takes place in varying contexts (Sloboda & Juslin, 2010). However, there is a lack of studies concerning the factors that influence contextual listening choices (Kamalzadeh, Baur, & Moller, 2012). Recent research implicates the importance of control and choice in everyday listening and suggests that Mehrabian and Russell's (1974) Pleasure-Arousal-Dominance framework might apply to everyday music listening. One way to examine this in the context of control and choice is by considering contextualized music playlists.

2. Aims: The present research considered everyday music listening in the context of eight situations, classified as high or low on Mehrabian and Russell's (1974) Pleasure-Arousal-Dominance (PAD) dimensions.

3. Method: 344 participants (43 x 8 situations) completed a questionnaire that included questions about themselves and a playlist task. Individuals were randomly given one of eight situations (e.g., a house party with friends, jogging with an mp3 player), for which they created a 10-12 song playlist and then answered questions about the songs they chose for that particular situation.

4. Results: The music selected by participants (as playlists) for the eight situations differed along two dimensions, namely arousing and sophisticated. Rather than selecting music that would moderate arousal (Berlyne, 1971), results indicated that participants employed an arousal-optimization strategy. There were also differences in the music selected across situations on the sophisticated dimension. The findings also suggest that music chosen for situations is subject to injunctive norms, such that there was considerable variation between the music chosen for listening to on public transport or when washing dishes, whereas music selected for a wedding was perceived as more homogenous.

5. Conclusions: While today's technology allows individuals to easily select music from large collections of familiar and preferred music, the present results demonstrate that music choice is not driven simply by personal preference. The differences in the songs selected for contextual listening are evident in terms of musical descriptors, BPM, and how similar individuals believe their playlist is to those made by other people for the same situation. More generally, the results indicate that Mehrabian and Russell's PAD dimensions offer a useful framework for considering the relationship between music and the environmental context in which it is experienced.

NEW MEDIA USE AND MUSIC APPRECIATION: CONNECTIONS BETWEEN CONTEMPORARY ACCESS TO MUSIC AND ITS PERCEIVED VALUE

Hauke Egermann[1], Daniel Kahlhoefer[1], Christof M. Schultz[1], Marten Seedorf[1]
[1]Audio Communication Group, Technische Universität Berlin, Germany
Time: 15:00

1. Background: In the course of a quick technological advancement and a wide dissemination of digital entertainment media, music has developed a certain ubiquity and free accessibility. Modern music streaming services offer millions of songs with little or no payment needed. Critical voices, especially in the musicological discourse, describe within this development a strong bias towards the trivialization of music and its consumption, towards the economical and ideational devaluation of music itself. So, within this context, one could ask, if new media use for music listening does affect music appreciation negatively.

2. Aims: First, by means of factor analysis, the complex construct appreciation of music will be analyzed and divided into sub-dimensions. We will then try to determine if these aspects are related to certain ways of mediatized music consumption. Furthermore, we will test also for specific connections between the use of music streaming services and a decreased appreciation of music.

3. Methods: In an online survey 125 participants rated agreement with 43 items describing their music appreciation in various ways. Subsequently, they were also asked about their music consumption habits (access technology and use frequency). The music appreciation items were then subject to a principal component analysis.

4. Results: The resulting components show several dimensions in which the elusive construct of music appreciation can be segmented: investment willingness/economical appreciation, cognitive appreciation, social importance, and cultural importance. Furthermore, those components are correlated with various ways of mediatized music consumption. However, the use of free streaming services does not negatively interfere with any dimension of music appreciation.

5. Conclusions: The interpretation of these results leads to questioning traditional forms of music appreciation and current discourses within cultural policy. Economical appreciation presents itself as a political issue that deals with the cultural and economic value of music in the digital age. So the idea of appreciation itself has to be discussed as an object of a constant social change. In principle, we believe that modern forms of music consumption do not contradict music appreciation, but they are in steady confrontation and exchange with it.

[1C] Music and Movement 1

MONTREAL 14:00-16:00 Monday 4 Aug 2014

DO LISTENER'S BODY MOVEMENTS INFLUENCE THEIR PERCEPTION OF RHYTHM?

sempre: Society for Education, Music and Psychology Research

Li-Ching Wang[1]
[1]Centre for Music and Science, University of Cambridge, United Kingdom
Time: 14:00

Listener's embodied responses to music have been widely explored recently in terms of music type, music environment, social function, and emotion. However, how body movement influences music perception has not been addressed in depth. This paper aims to explore whether or not body movement influences the perception of rhythmic complexity and memory of rhythms. Two experiments were conducted using rhythms categorised in terms of rhythm complexity according to Shmulevich & Povel (2000), at a tempo of 126 BPM. In experiment 1, 30 participants judged rhythm complexity by giving a score ranging from 1 (simple) to 5 (complex) in the following 5 conditions in random order: no-move, move-head, move-hand, move-feet and move-whatever they want when listening to the rhythms. A repeated measures mixed ANOVA revealed a main effect of rhythmic structure ($p=.001 <.05$) indicating participants could tell the difference between simple and complex rhythms. There is an interaction effect of movement and participant's preference of movement ($p=.011 <.05$). Test of between subjects effects revealed that enjoyment was also significant ($p=.043 <.05$). Bivariate (Pearson) Correlation shows the correlation coefficient between complexity scores given by participants and those calculated in Schmulevich & Povel's (2000) paper in every task is significant at the 0.01 level (2 tailed) while Move-head task has the highest correlation coefficient. In experiment 2, 17 participants were asked to answer Yes or No in response to rhythm sets in the form of A-B-X under no-move and move-head conditions. "Yes" means X is either A or B while "No" means X is neither A nor B. A repeated measures mixed ANOVA revealed a significant effect on the interaction between rhythm structure and movement type ($r=.024 <.05$). Bivariate (Pearson) Correlation shows the correlation coefficient between enjoyment and memory performance is again significant ($p<.001$). The results from two experiments suggest that body movements may not necessarily render the rhythms more easily perceivable or memorable, but that they provide an opportunity for listeners to utilise music in different ways. An individual's preferences as to movement, and their enjoyment when moving, can affect their perception of rhythm complexity and memory of rhythm.

DOES WATCHING BODY MOVEMENT AFFECT RHYTHM PERCEPTION?

Kyung Myun Lee[1]
[1]*Graduate School of Convergence Science and Technology, Seoul National University, Korea, Republic of*
Time: 14:30

Rhythm is closely related to body movement. Previous studies have shown that the way people move their bodies to music affect their auditory perception of the rhythm structure (Philips-Silver &Trainor, 2005; 2007). If so, can just watching body movement influence rhythm perception? To answer this question, we designed a behavioral experiment, in which participants were asked to detect timbre changes while watching dance videos. Each stimulus was an isochronous rhythm and the task was to press a button when the timbre of the rhythm changed from the woodblock into the guitar. The changing point was randomly assigned and counterbalanced. During the task, participants watched a video of dancing to a quadruple meter or of a circle picture changing its color every fourth beat. In the control condition, no video was provided. Two video conditions showed response time (RT) differences among beat positions, whereas the control condition did not. This result suggests that the visual metrical cue could influence the rhythm perception.

I LIKE TO MOVE IT: HOW MUSIC-INDUCED MOVEMENT RELATES TO MUSIC PREFERENCE, FAMILIARITY, AND MOVABILITY

Birgitta Burger[1]*, Geoff Luck*[1]*, Marc Thompson*[1]*, Suvi Saarikallio*[1]*, Petri Toiviainen*[1]
[1]*Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland*
Time: 15:00

Listening to music makes us move. Music-intrinsic features, for instance pulse clarity, or emotional content of music, have been shown to influence music-induced movements, as well as individual factors, such as personality. Additionally, music-induced movement may be shaped by musical preference, familiarity with the music, and its movability. This study aims at investigating relationships between musical preference, familiarity, and movability and music-induced movement. It was hypothesized that there is a U-shaped relationship for preference (i.e., participants move more both when they like and when they dislike the music) and linear relationships for familiarity and movability (i.e., participants move more the more familiar the music is and the easier it is for them to move along with the music). Sixty participants were presented with 30 musical stimuli representing different genres of popular music. Participants were asked to move along in a natural way. After each stimulus, participants gave ratings for preference, familiarity, and movability on a scale from 1 to 5. An optical motion capture system was used to record participants' movements, from which 17 movement features related to speed, acceleration, complexity, amount, and rotation were subsequently extracted. Fifteen out of the 17 movement variables showed a U- / J-shaped relationship with the preference ratings, indicating that participants had higher movement feature values with either low or high than with moderate preference for the music. For familiarity, 13 movement features demonstrated highest values when participants were rather familiar with the music (rating score 4). 16 movement features exhibited linearly increasing relationships with the movability ratings, suggesting that participants moved more the easier they found it was to move to the music. The study revealed relationships between music-induced movements and musical preference, familiarity, and movability. The movability results seem straightforward: people moved more when they found it easy to move to the music. Moreover, the familiarity results suggest that music triggers most movement when participants are somewhat familiar, but not too familiar. Finally, preference seems to influence movement in a bidirectional way, which could be related to being emotionally involved both when liking and disliking the music.

TEMPO IN BAROQUE MUSIC AND DANCE

Esther Coorevits[1]*, Dirk Moelants*[1]
[1]*Ipem, Dept. of Musicology, Ghent University, Belgium*
Time: 15:30

ABSTRACT

Recently, there is a growing interest in studies on the relationship between music and movement. Insight in the relation between dance and music is particularly important for the Baroque period, as music and dance were directly related, even if music was not used to dance to. In Baroque dance, particular dance steps and the character of different dance types demand a specific tempo. However, in musical performance practice, the tempo variation can be very large and the link with the original dance movement is often lost. The aim of this study is to compare the interpretations of dancers and musicians in an experimental setting. The study consists of two parts. First, we investigate the influence of dance movement on the interpretation of a series of dances by musicians. The pieces were recorded with and without dance and we compare tempo and timing in the different versions. In the second part, dancers performed a particular choreography on music that varied in tempo. Video analysis and questionnaires were used to evaluate the different performances. These results were compared with the tempo of musical recordings of similar dance types. Results show a clear difference between music and dance performance. Musicians adapt their interpretation when performing together with the dancers and the optimal tempo zone found for certain Baroque dances coincides only partly with the tempi commonly found in music recordings. The direct link between music and movement and its mutual influence illustrates the importance of an embodied approach in music performance, where in this case dance movement gives concrete information for a 'historically informed' performance.

1. INTRODUCTION

Regardless of style or region of origin, music and dance are closely related art forms. Not only is dance almost always accompanied by music, also do the movements associated with certain musical genres influence the way we interpret and perceive the music. Up to

now, scholarly research has focused on either dance or music, mainly based on written sources. Few attempts have been done to shed light on the mutual interaction between dance and music in an actual performance context.

Insight in the relation between dance and music is particularly important for the Baroque period, and this for a number of reasons. First, the direct link between dance practice and music is very important in this era. An important share of the music was composed with the intention to serve as dance music, both in a social context and in a theatrical setting. Second, the influence of dance music is very prominent in other genres: instrumental music often takes dance movements as a model and also in vocal genres as cantatas or oratorios references to dance movements are prominent. Therefore the music of the Baroque era cannot be fully understood without insight in the dance practice of that period.

In this paper we present two empirical studies that explore the interaction between dance performance and the interpretation of Baroque music. Empirical studies that deal with historical performance practice are rare. A few studies have focused on specific aspects of timing such as overdotting (Fabian & Schubert, 2004) or 'notes inégales' (Moelants, 2011), comparing the interpretation of different performers. But the relation with dance movement is not yet taken into account. At the same time, we see a growing interest in studies on the relationship between music and movement in more general terms. The topic appears in different areas, such as musical development (Phillips-Silver & Trainor, 2005), music education (Campbell, 1991), rehabilitation (Thaut & Abiru, 2010), musical communication (Moelants, Demey, Grachten, Wu, & Leman, 2012) or the influence of music on movement (Leman et al., 2013). These studies have given rise to a large spectrum of methods that can be applied to study the relationship between dance and music (Leman & Naveda, 2010; Maduell & Wing, 2007). But none of these has placed their research in a historical context, looking at the relation between historical dance and music performance. On the other hand several studies have stressed the importance of dance for the interpretation of Baroque music but did not look at an actual performance (Franko, 2011; Little & Jenne, 2001). In this paper we will first go deeper into the relation between Baroque dance and music, give an overview of some aspects of dance which have a crucial influence on the interpretation of the music and present two experiments in which the relation between music and dance is explored in a real performance context.

2. BACKGROUND

2.1. Dance and music in the Baroque era

During the 17th century, dance reached its status as a serious art discipline in the Western tradition, which, eventually, would lead to the 'romantic' ballet of the 19th century (Franko, 2011). Leading nation in the development of dance as an art discipline was France. Court dancing, refined and graceful in style and technique, symbolized French culture and was considered superior to any other dance form in Europe (Little & Jenne, 2001). The dissemination of French culture through noble dance was not the least accomplished by French dancing masters who were employed at every notable European court. To quote Pierre Rameau (1725): "There isn't any European court that hasn't employed a French dancing master." Along with the dance, they spread the music accompanying it. Thus, both the European aristocracy and court musicians and composers came into contact with French dances and their music. Music theorists of the 17th century encouraged musicians to refine themselves in the art of dancing in order to feel the natural rhythmical flow and pulse of the music (Muffat, 1698).

The chief tools in the dissemination of court dance were the many dance treatises and manuals that were published and spread throughout Europe. In 1700, a dance notation system developed by Pierre Beauchamp was published by Feuillet in his Chorégraphie, ou l'art de décrire la danse (1700). It elaborated an extended and complex dance vocabulary and was the most popular system at that time. Its chief components are symbols to show the movements of the dancers feet and legs, the distribution of these symbols along the dance path where the bars of music are indicated and the corresponding musical tune on top of page that shows how music and dance are related (Pierce, 1998).

The vision of the dancer-choreographers was that writing down dances in a fixed system could ensure the correct and complete transferal of dances from one generation to another (Franko, 2011). The codified steps, path of choreography and its relation to the music do give us a lot of information to apprehend court dance. But at the same time, the complexity can obscure the correct interpretation. Dance masters were aware of this and wrote treatises about the interpretation of dance-notation. Famous examples include Traité de la Cadence (Feuillet & Pécour, 1704) and Le Maitre à Danser (Rameau, 1725). Two scholars who made a major contribution to research on the relation between Baroque music and dance are Hilton (Hilton, 1977, 1981, 1986) and Little (Little, 1975a, 1975b; Little & Jenne, 2001).

But the influence of the dance goes beyond actual dance music. During the 17th century, stylized dance music emerged that contained more elaborated melodic and rhythmic traits, intended for concerts without actual dance performance (Spitzer & Zaslaw, 2005). One of the questions here is whether the original dance practice is still traceable in the performance of these works. Despite the absence of choreography, the two were not entirely disconnected. Both musicians and composers had practical knowledge of ballroom and theatre dance, which inevitably influenced their compositions and performances. J.S. Bach, for example, wrote a large amount of music of which not only the title but also the musical traits imply a connection to French court dancing. On the other hand, very popular dances in instrumental music (e.g. the allemande) were not performed as choreography anymore in the 18th century.

This opens a lot of questions on interpretation, both for dancers as musicians interested in historical performance practice. One of the main problems when dealing with the interpretation of Baroque dance music is the choice of the tempo.

2.2. Tempo and its relationship to choreography

Tempo-indications in Baroque music are mostly expressed verbally and by consequence open to interpretation. They can refer to the character, movement or the expression in which the music should be played or performed, for example vivement or gay, andante or allegro (Cyr, 1998). Apart from verbal indications, the time signature or meter of a musical piece bears a connotation of tempo too. In general, the speed of the counting unit (the denominator of the fraction) slows down when the unit gets smaller. A bar in 3/2 is then supposed to be played slower than a bar in 3/4. This way, tempo and meter are closely related to each other, but not equal. When reading tempo-indications, a pitfall is the metric level of the meter the indication is pointing to. A dance could be interpreted slowly at the level of the half note, but livelier on the level of the quarter note. By exploring the metric structure it is possible to find some indications for the tempo and character of the music (Little & Jenne, 2001). In any case, playing music in the appropriate tempo was considered essential to good musical taste, as Étienne Loulié points out in his Éléments (1696):

"Mais je me flatte que ceux qui ont le goust fin & qui ont éprouvé combien un Air perd de sa beauté lorsqu'il est executé trop viste ou trop lentement, me sçauront bongré de leur donner un moyen (…) pour en connoître le veritable mouvement (…)"

In dance music, the correct tempo is essential for the character of the dance, the corresponding choreography and the distinction of different dance genres (Donington, 1989). French music theorists were aware of this problem at that time, and developed an instrument to define tempo-indications more exactly, relaying on the movement of a pendulum (de La Chapelle, 1737; l'Affilard, 1705; Loulié, 1696; Meyerhöfer, 1996; Pajot, 1735). Pendulum data were preserved for a small selection of dances, but they are not as unequivocal as our present metronome numbers and the correct interpretation of these data is still under discussion (Harris-Warrick, 1992; Heijdemann, 1984; Hilton, 1981; Jerold, 2010; Kroemer, 2001; Van Biezen, 1984). This is where actual dance practice can give insight in the difficult question of tempo in Baroque dance.

First of all, the tempo of the music should be appropriate for the execution of dance steps, especially the leaps. These can be performed slow or fast, but it is physically impossible to remain suspended in the air infinitely. Also, one measure of music may contain many steps, which would take a dancer more time and thus a slower tempo. In general, Baroque court dance should be performed with an air of effortlessness and elegance, and all dance steps and movements should be fluently merged into each other. A skilled dancer will be able to perform the dance steps more quickly without getting untidy and will not loose the smoothness in performing steps very slowly (Little, 1975b). Also the character of the dance is important for the tempo: choreographies containing sliding steps and few hops bear a more dignified, serious character and therefore will be performed more slowly than choreographies containing many leaps and jumps. The way a specific step is executed or articulated, for example a deep and slow opposite to a fast and crispy plié (bending of the knees), tells a lot about the character of the dance and depends largely on the tempo of the music (Little, 1975a). In the research on Baroque dance music, some scholars already took the notated choreography into consideration (Cobau, 1984; Harris-Warrick, 1992; Hilton, 1977; Little, 1975a; Little & Jenne, 2001; Qureshi, 1994), but up to now, systematic research on actual dance practice is lacking. In this paper, two studies are presented which explore the relationship between tempo, musical practice and dance performance.

3. EXPERIMENT 1: THE INFLUENCE OF DANCE ON MUSIC PERFORMANCE

3.1. Procedure

The choreography used for this experiment is "The Submission", a ballroom dance created in 1717 by the English dancing master Kellom Tomlinson on music by John Loeillet. It consists of three musical parts, each with its own character, on which one choreography is danced. A slow movement in 3/4 is followed by a lively menuet, also in 3/4, while the last movement is a fast rigaudon in 4/4. Also, the character of the choreography changes with the dance types: The menuet contains the typical *pas de menuet* and the faster movements contain more lively steps.

A couple of experienced baroque dancers performed with a musical ensemble consisting of a violin, a viola, a viol and a harpsichord. Initially, dancers and musicians rehearsed separately. As such, they fixed their own ideas on interpretation of the dance suite. At the end of the rehearsal, the musical performance was recorded, with the musicians not knowing the choreography. After this, the dancers joined the ensemble. First musicians and dancers performed the dance together, without communicating in advance. After this first joint performance, the dancers gave feedback and, together with the musicians, tried to come to an 'ideal' performance in which everybody agreed on the interpretation. Individual recordings of the musicians were made using contact microphones. Also the complete ensemble-sound was recorded and a video recording was made for control.

3.2. Analysis

Four versions of the piece were recorded: the instrumental version without dance and three versions with dance (one without and two with feedback). As we are mainly interested in the timing and tempo of the whole, we decided to base our analysis on the viol part as it plays continuously and articulates the meter most clearly. To determine the length of each event, we used a manual onset detection system using the program 'praat', which gives an automatic analysis of pitch and intensity as well as a spectrogram visualization, with parallel audio feedback. The use of this program makes it relatively easy to locate the onsets with a precision of a few milliseconds.

3.3. Results

The results of the timing analysis are summarized in fig.1. In the graph we clearly see the succession of three different parts. At the same time we see clear differences between the four versions. Three main changes were found in the analysis: a main change in tempo, differences in the endings of the first two parts and a change in the variability between the different versions.

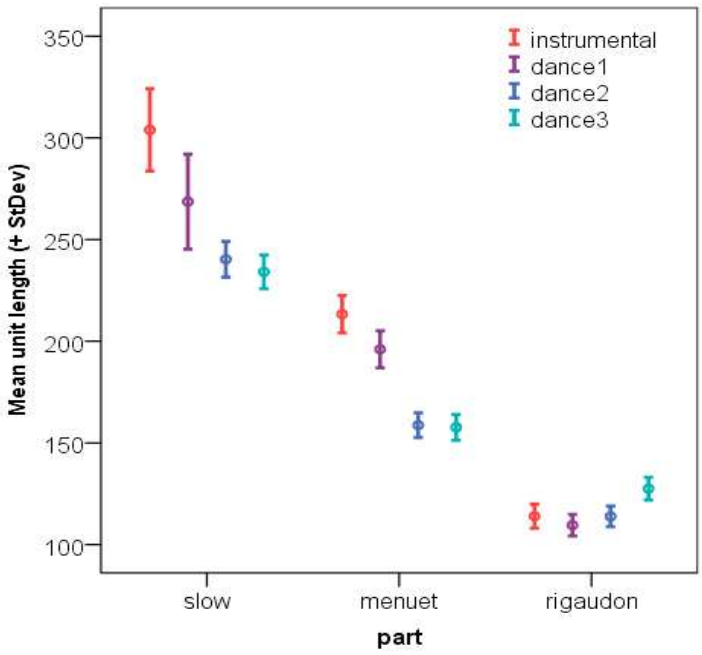


Figure 1: Timing for the viola da gamba in the four recordings of the three sections of "The Submission"

Is it clear that the musicians immediately speed up their tempo in the first two movements when the dance is added. Even without oral communication (dance1) musicians feel that the dance demands a particular tempo. After communicating with the dancers, the tempo is adapted even more and the first two parts are performed even faster (dance2). After the second session the dancers were happy with the tempo for the first two parts, however they thought the final part was performed too fast. Therefore a final version was recorded (dance3), where we see that the timing for the first two parts is very similar to the previous version, while the last part is played consistently slower. The second change considers the final retard in the first two parts. When recording the music without dance, the musicians tended to take a break between the parts. In the first version including dance we see that this largely remains, as the musicians stopped and waited for the dancers to give a new start. However, the dancers perceive the choreography as a whole and need fluent transitions between the different parts, as the dance movement goes on.

A final effect is the lowering of the variance in the versions with dance, notably in the first two parts. Figure 1 shows the mean length of the unit in the three different parts for the four versions. Also in this case we can see a different attitude between dancers and musicians. Whereas musicians tend to exaggerate phrasing and rhythmic contrast, dancers actually prefer a more stable performance in which they can keep the cadence of their movement.

4. EXPERIMENT 2: COURANTE AND SARABANDE IN MUSIC AND DANCE

4.1. Procedure

The second experiment focuses on two popular dance types, the courante and the sarabande. The study consists of two parts: an analysis of tempo in existing performances and a study of actual dance performances with music of varying tempo.

First, the tempo of recorded sarabandes and courantes by French composers was analyzed. Pieces were collected using Spotify, based on a selection of representative composers of which the scores were available. The selection also contained different recordings of the same dance, if played by different performers, but only recordings on period instruments were retained. Based on the original time signature of each piece, the tempo was analyzed manually by tapping along with the music.

In the second part, we looked at choreographic performance to explore how dance movement is influenced by tempo. Based on surviving choreographies, a basic choreography for a courante and a sarabande were designed, containing the characteristic steps of these dances. To optimize the choreographies, the result was discussed with four dancing masters. For the courante, the melody of the dance suite 'La Duchesse' (*La Duchesse*, 1701) was used, for the sarabande, the music of 'La Royalle' (Pécour, 1712) was adopted. For both pieces an orchestration was made based on historical performance practice. These versions were transformed into MIDI files and rendered in a variety of tempi, starting from 80 BPM and moving up and down based on a division of the tempo-octave in 6 equal steps using a log-2 scale. This resulted in the tempi seen in fig. 2, which were applied to the quarter note of the melody of the sarabande, and the half note of the courante.

16 dancers participated in the experiment. They were asked to perform the choreography of the sarabande and courante on the music presented to them. The dancers performed the choreography twice in each tempo, which was recorded on video. After the two performances, they were asked to evaluate the tempo of the music on the basis of 1) the character or the dance, 2) fluency of their dance

movements, 3) how they felt rhythmical cadence, 4) the difficulty to perform the choreography, 5) how comfortable they thought the tempo was on a 7-point Likert scale. Starting with the 'neutral' tempo of 80 BPM, alternately, a faster and a slower tempo were played until the dancers indicated that it was impossible to perform the choreography.

4.2. Analysis

Analysis of music recordings

944 recordings of music by Lully, Louis and François Couperin, Marais, Hotteterre, d'Anglebert, Chambonnières, Rameau and Jacquet de la Guerre were used to give a representative view of tempi in historically informed performances of French baroque music. The collection consisted of 507 sarabandes (mostly in 3/4, with some examples in 3/2 or 6/4) and 437 courantes (mostly in 3/2, with some examples in 3/4). Whereas the tempi of the sarabandes were very similar despite the metrical difference, the group of courantes in 3/4 were clearly performed faster. As this meter does not correspond to the meter used in the dance study, 26 courantes were excluded from the analysis.

The distributions of tempi for both dances are very similar, with means of 68.86 and 68.47 bpm for courante and sarabande respectively. The range of tempi is larger for the sarabande, which is reflected by a higher standard deviation (18.7 for the sarabande and 10.2 for the courante). The distribution is also more skewed, which indicates that there is quite a large group of slower sarabandes, but this is counterbalanced by a small group of fast pieces.

Analysis of dance performance

The analysis of the video recordings of the dance performances was done manually and comprised three criteria: the synchronization of the dancers to the music, their precision and smoothness. For the sarabande, every performance got a score on 16, while the courante was assigned a score of 12, in accordance with the number of bars in the dance.

The results of the video analysis show that precision and smoothness score the highest for 80 bpm with the courante (mean=88.9%), and with the sarabande for 113 bpm (mean=80.8%). The peak is much sharper for the courante, the sarabande has high scores for al the tempi between 80 and 127 bpm. De dancers synchronize best at 90 bpm with the courante (mean=88.9%) and at 101 bpm with the sarabande (mean=82.9%). Also, nobody was able to perform the sarabande slower then 40 bpm and the courante faster than 180 bpm. The dancers themselves rated 80 bpm as the optimal tempo for the courante, and 90 bpm as optimal for the sarabande.

The results are summarized in figure 2, where we see the distribution of tempi found in the music recordings (grey bars), the average score of the evaluations by the dancers (experience, black line) and the average score of the smoothness and precision parameters from the video analysis (perception, grey line).

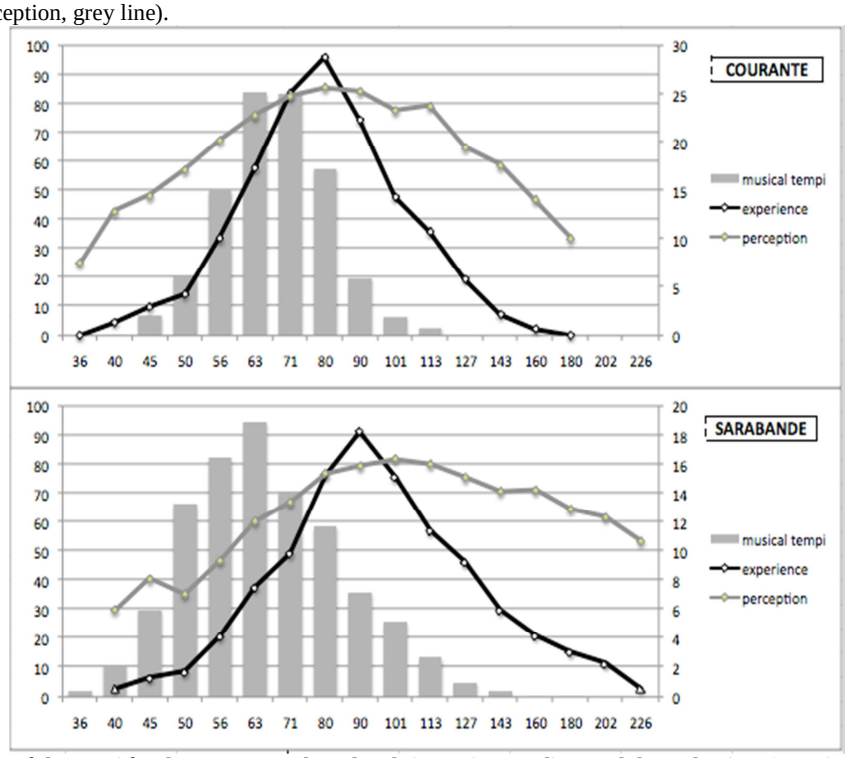


Figure 2: Comparison of the tempi for the courante and sarabande in music recordings and the evaluations (experience of the dancers and perceptual analysis) of the dance performance

5. DISCUSSION AND CONCLUSIONS

The results of the first experiment confirm that dance movement can have an actual influence on the interpretation of dance music by musicians. Verbal communication between dancers and musicians can help to generate a musical performance that fits the choreography best, but even when bodily communication alone is used during performance, musicians experience an influence of the dance movement on their interpretation, which mainly effects their tempo and timing.

However, lots of musicians are not aware of the corporeal aspects of Baroque dance music and hence an essential part of this music is lost. The results of the second experiment show that the interpretation of the sarabande by musicians is in general slower than the tempo dancers prefer. A similar but less pronounced trend is observed for the interpretation of the courante. Moreover, the relationship of tempo between courante and sarabande as defined by the dance experiment is in accordance with the proportions of the time signatures 3/4 en 3/2 and the characteristics of these two dance types as described in historical sources of the Baroque period (d'Alembert, 1752; de Brossard, 1703; de Saint-Lambert, 1702; Freillon-Poncein, 1700; Masson, 1705; Montéclair, 1736; Muffat, 1698; Rameau, 1725; Rousseau, 1768). In the research field of historical music performance, the information gathered from dance and choreography could be of great importance for the realization of a historically correct interpretation.

On the other hand, the question still remains how strictly the character of the dance was maintained when there was no active dance practice involved. As already mentioned, court composers were aware of the aspects of dance performance, also when they composed concert music, but an evolution to more complex forms inevitable lead to compositional and interpretational freedom. Also, not all dances were commonly danced during the Baroque, which opens the question if dance movement did influence composition and performance in such cases. Nevertheless, collaborations between dancers and musicians can lead to new and fruitful insights in the puzzling language of Baroque Dance.

6. ACKNOWLEGDEMENT

We would like to thank Lieven Baert and Josephine Schreibers the two dancers involved in the first experiment, as well as the musicians Stefaan Smagghe and Kersten Cottyn.

Special thanks goes also to Guillaume Jablonka, Béatrice Massin, Dorothée Wortelboer and Johan Goessens, the four dancing Master who thought us a lot and helped us with the choreography and the second experiment.

7. REFERENCES

Campbell, P.S. (1991). Rhythmic Movement and Public School Music Education: Conservative and Progressive Views of the Formative Years. Journal of Research in Music Education, 39*(1), 12-22.*

Cobau, J. (1984). The Preferred Pas de Menuet. Dance Research Journal, 16*(2), 13-17.*

Cyr, M. (1998). Performing Baroque *music (R. G. Pauly Ed.). Portland: Amadeus press.*

d'Alembert, J.l.R. (1752). Éléments de Musique théorique et pratique, suivant les Principes de M. Rameau. *Paris: Jombert.*

de Brossard, S. (Ed.) (1703) Dictionaire [sic.] de Musique (1st edition ed.). Paris: Ballard, Christophe.

de La Chapelle, J.-A. (1737). Les Vrais principes de la musique. *Paris: Veuve Boivin.*

de Saint-Lambert, M. (1702). Les Principes du clavecin contenant une explication exacte de tout ce qui concerne la tabulature & le clavier. *Paris: Ballard, Christophe.*

Donington, R. (1989). The interpretation of early music: New revised edition. *London: Faber and Faber.*

Fabian, D., & Schubert, E. (2004). Expressive Devices and Perceived Musical Character in 34 Performances of Variation 7 from Bach's Goldbergvariations. Musicae Scientiae, 7*(1 suppl), 49-71.*

Feuillet, R.A. (1700). Choregraphie: ou L'art de décrire la dance. Paris: Chez l'auteur et chez Michel Brunet.

Feuillet, R.A., & Pécour, L.-G. (1704). Traité de la cadance. Recüeil de dances. *Paris: Chez le sieur Feüillet.*

Franko, M. (2011). Writing For The Body: Notation, Reconstruction, and Reinvention in Dance. Common Knowledge, 17*(2), 321-334.*

Freillon-Poncein, J.-P. (1700). La véritable manière à jouer en perfection du haut-bois. *Paris: J.Collombat.*

Harris-Warrick, R. (1992). Interpretation of Pendulum Markings for 18th-Century French Dance. *Paper presented at the The marriage of music and dance: National Early Music Association.*

Heijdemann, F. (1984). De tempo-aanduidingen bij L'Affilard en Pajot (d'Onzembray). In K. Vellekoop (Ed.), Tempo in de Achttiende Eeuw *(pp. 26-36). Utrecht.*

Hilton, W. (1977). A Dance for Kings: The 17th-Century French "Courante". Its Character, Step-Patterns, Metric and Proportional Foundations. Early Music, 5*(2), 161-172.*

Hilton, W. (1981). Dance and Music of Court and Theater: Selected Writings of Wendy Hilton*: Pendragon Press.*

Hilton, W. (1986). Dances to Music by Jean-Baptiste Lully. Early Music, 14*(1), 51-63.*

Jerold, B. (2010). The French Time Devices Revisited. Dutch journal of music theory, 15*(3), 169-189.*

Kroemer, J. (2001). Le cronometre de Monsieur Feuillet. *(PhD Dissertation), University of Vienna.*

l'Affilard, M. (1705). Principes très faciles pour bien apprendre la musique, 5th edition. *Paris: Christophe Ballard.*

La Duchesse. *(1701). Chorégraphie, par DESCAN, (14884). Bibliothèque nationale de France, Département des manuscrits.*

Leman, M., Moelants, D., Varewyck, M., Styns, F., van Noorden, L., & Martens, J.-P. (2013). Activating and Relaxing Music Entrains the Speed of Beat Synchronized Walking. PLoS ONE, 8*(7), e67932.*

Leman, M., & Naveda, L. (2010). Basic Gestures as Spatiotemporal Reference Frames for Repetitive Dance/Music Patterns in Samba and Charleston. Music Perception: An Interdisciplinary Journal, 28*(1), 71-91.*

Little, M.E. (1975a). The Contribution of Dance Steps to Musical Analysis and Performance: "La Bourgogne". Journal of the American Musicological Society, 28*(1), 112-124.*
Little, M.E. (1975b). Dance under Louis XIV and XV: Some Implications for the Musician. Early Music, 3*(4), 331-340.*
Little, M.E., & Jenne, N. (2001). Dance and the Music of J.S. Bach*: Indiana University Press.*
Loulié, É. (1696). Elements ou Principes de musique mis dans un nouvel ordre*. Paris: Christophe Balard.*
Maduell, M., & Wing, A.M. (2007). The dynamics of ensemble: the case for flamenco. Psychology of Music, 35*(4), 591-627.*
Masson, C. (1705). Nouveau traité des règles pour la composition de la musique*. Paris: Ballard, Christophe.*
Meyerhöfer, D. (1996). Lichtenberg und die Sammlung wissentschaftlicher Geräte des Frankfurter Patriziers Johann Friedrich Armand von Uffenbach Lichtenberg Jahrbuch 1995*. Saarbrücken.*
Moelants, D. (2011). The Performance of Notes Inégales: The Influence of Tempo, Musical Structure, and Individual Performance Style on Expressive Timing. Music Perception: An Interdisciplinary Journal, 28*(5), 449-460.*
Moelants, D., Demey, M., Grachten, M., Wu, C.-F., & Leman, M. (2012). The Influence of an Audience on Performers: A Comparison Between Rehearsal and Concert Using Audio, Video and Movement Data. Journal of New Music Research, 41*(1), 67-78.*
Montéclair, M.P. (1736). Principes de musique*. Paris: veuve Boivin.*
Muffat, G. (1698). Florilegium Secundum*. Denkmäler der Tonkunst in Öterreich, (4). Vienna.*
Pajot, L.-L. (1735). Description et usage d'un métromètre, ou machine pour battre les mesures & les temps de toutes sortes d'airs. Histoire de l'Académie Royale des Sciences, 1732*, 182-196.*
Pécour, G.-L. (1712). La Royalle Nouveau recüeil de dance de bal et celle de ballet, contenant un très grand nombres ["sic"] des meillieures ["sic"] entrées de ballet *(pp. 1-5). Paris: Gaudrau, M.*
Phillips-Silver, J., & Trainor, L.J. (2005). Feeling the Beat: Movement Influences Infant Rhythm Perception. Science, 308*(5727), 1430.*
Pierce, K. (1998). Dance Notation Systems in Late 17th-Century France. Early Music, 26*(2), 287-299.*
Qureshi, R.J. (1994). The influence of Baroque dance in the performance of Johann Sebastian Bach's Six Suites a Violoncello Senza Basso. *(Doctoral Thesis), Rice University.*
Rameau, P. (1725). Le Maître a danser. *Paris: J. Villette.*
Rousseau, J.-J. (1768). Dictionnaire de musique*. Paris: la venue Duchesse.*
Spitzer, J., & Zaslaw, N. (2005). The Birth of the Orchestra : History of an Institution, 1650-1815*: Oxford University Press, USA.*
Thaut, M.H., & Abiru, M. (2010). Rhythmic Auditory Stimulation in Rehabilitation of Movement Disorders: A Review Of Current Research. Music Perception: An Interdisciplinary Journal, 27*(4), 263-269.*
Van Biezen, J. (1984). Het tempo van de Franse Barokdansen. In K. Vellekoop (Ed.), Tempo in de Achttiende Eeuw *(pp. 7-25). Utrecht.*

[1D] Aesthetic Perception and Emotions

LIÈGE 14:00-16:00 Monday 4 Aug 2014

MUSICAL PREFERENCES OF MUSICOLOGY STUDENTS

Paul Elvers[1], Wolfgang Fuhrmann[2], Timo Fischinger[1]
[1]Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany, [2]Institute of Musicology and Media Science, Humboldt University Berlin, Germany

Time: 14:00

1. Background: An increasing amount of research on musical preferences has been done including diverse topics such as the development of musical preferences (Bonneville-Roussy, Rentfrow, Xu, & Potter, 2013) underlying dimensions and connections to personality traits (Rentfrow & Gosling, 2003) social aspects (Abrams, 2009; North & Hargreaves, 2007) or listening behaviour (Dunn, de Ruyter, & Bouwhuis, 2011). The outcome of these works indicate that musical preference is a multidimensional phenomenon, influenced by various factors such as age, personality, socialization and music specific features.

2. Aims: Compared to the research on other aspects of musical preferences, to the authors knowledge there has been no substantial attempt to investigate musical preferences among musicologists. Though surveying them could be highly rewarding: Because of their prolific ability of musical judgement, they are supposed to serve as experts of musical preferences. The study seeks to examine the following questions: What kind of music do musicology scholars listen to? What aspects influence their musical preferences? How is this related to the special case of studying musicology?

3. Method: First, 30 undergraduate musicology students where asked to form pairs in which each person interviews the other in order to write a short essay about each other's musical preferences. Second, approximately 400 musicology scholars from two different universities were asked to complete an online survey, which was based on the results of the essays.

4. Results: A first analysis of the essays shows that almost all of the students have musically active families and started musical practice in early childhood. Many of them developed preferences for classical music in this time. In early adolescence preferences shift to popular music genres, where peer and media influences become predominant. These shifts are often linked with strong musical "key experiences".

5. Conclusions: The preliminary results from the essays indicate that musical preferences of musicology students are characterized by a great openness for different styles and genres, resulting from an intensive musical affiliation during childhood and adolescence. However, the results are preliminary. A fuller account of the research will be available at the end of January.

Note: List of references upon request.

PERCEPTION OF METRIC TIMING PATTERNS IN MALIAN JEMBE-MUSIC

Hans Neuhoff[1], Rainer Polak[1], Timo Fischinger[2]
[1]Cologne University of Music and Dance, Germany, [2]Max Planck Institute for Empirical Aesthetics, Germany
Time: 14:30

1. Background: Chronometric analyses of Malian jembe-music by Polak (2010) suggest that its characteristic feel rests upon non-isochronous subdivisions of the beat. Each feel is marked by a specific collocation of two or three different sub-divisional units (pulses): "short", "medium," or "long" as evidenced by statistical analyses of the performance data. Following London (2004) and Kvifte (2007), a specific collocation of pulse classes, like short-medium-long (SML), is called a metric timing pattern. London (2010), however, questions the existence of more than two different pulse classes in beat subdivision on psychological and theoretical grounds.

2. Aims: In order to shed light on this issue, we tested both assumptions by employing a pairwise comparison paradigm, presenting pairs of timing-manipulated versions of ensemble phrases from a well-known piece of Jembe-music, called Manjanin.

3. Method: 24 professional Malian musicians participated in our study (Bamako, December 2012). A 4-beat-cycle with ternary subdivision was used, displaying a dominant SML timing pattern. Stimuli consisted of pairs of systematically varied versions of the same phrase, including the original SML pattern. Participants were asked to rate (1) if the two items of each pair were the same or different, and (2) on difference detection, which of the two was better Manjanin.

4. Results: It could be shown that, perceptually and cognitively, a non-isochronous short-long-long (SLL) metric timing pattern is constitutive for the Manjanin feel. Whereas the SML pattern, as identified by Polak (2010) by means of chronometrical analyses of the surface structures of Manjanin performances, pertains to production only.

5. Conclusions: Metric timing patterns appear to be perceptually and cognitively less fine-grained than the performance data suggest. The number of categorically different subdivisional units, or (sub)pulse classes, is perceptually and cognitively limited to two, viz short and long.

EMOTIONAL EXPRESSION THROUGH MUSIC IN 3-5-YEAR-OLDS

Suvi Saarikallio[1], Antti Yrtti[1], Minna Huotilainen[2]
[1]University of Jyväskylä, Finland, [2]University of Helsinki, Finland
Time: 15:00

1. Background: Children can both perceive and express basic emotions through music. Perception of happiness and sadness precedes other emotions, but already 4-year-olds correctly identify also anger and fear. Certain performance features (e.g. tempo, loudness) are employed for decoding emotions earlier than other, culturally learned features (e.g. mode). However, as study materials and procedures vary, little is known about the particular parameters that children at different ages employ to express emotions. Also, majority of studies have focused on emotion perception instead of expression.

2. Aims: We investigated three parameters – tempo, loudness, and pitch – in expressing three emotions – happiness, sadness, and ager – in pre-school children. We tested whether children would use all the parameters in distinguishing the emotions from each other, and whether this would differ between 3 and 5-year-olds. Differences based on gender and music background were also studied.

3. Method: Participants (N=37) included eighteen 3-year-olds (8 female, mean age 3.7) and nineteen 5-year-olds (10 female, mean age 5.2). A novel interface, called Music Boxes, was used to modify tempo, loudness and pitch with sliders in real time. The stimuli included thee children's songs, randomly ordered, and the task was to make each song sound as happy, sad, and angry as possible. Parents filled in questionnaires about their child's music training.

4. Results: Significant differences based on emotion were found for all parameters. Sadness received significantly lower values than happiness and anger for tempo, loudness, and pitch, while anger only differed from happiness by higher volume. The expressions of 3-year-olds and 5-year-olds did not significantly differ, except for 3-year-olds expressing anger with higher volume. No gender differences were observed, but children with musical training expressed anger with significantly lower pitch values than children with no training.

5. Conclusion: Overall, 3-year-olds appeared equally competent than 5-year-olds in separating sadness, happiness, and anger using tempo, loudness and pitch. Happiness and anger were clearly separated from sadness but not from each other, implying to predominance of arousal over valence in expression. Observed differences based on musical training encourage future research on music training's effects on emotional expression.

A MODEL OF PERSONAL MUSIC LISTENING: EMOTIONAL OUTCOMES WITHIN CONTEXTS AND BETWEEN LISTENERS

William Randall[1], Nikki Rickard[1]
[1]Monash University, Australia
Time: 15:30

Personal music listening has become central to everyday music use, with mobile technology defining a new era of music consumption. This portable and flexible style of listening allows for the immediate selection of music to fulfil emotional needs, presenting it as a powerful resource for emotion regulation. The experience sampling method (ESM) is ideal for observing personal music listening, as it assesses current subjective experience during natural everyday music episodes. The current study aimed to develop a comprehensive model of personal music listening, and to determine the interaction of variables that produce either hedonic benefit or detriment. Data

were collected from 195 participants using the MuPsych app, a mobile ESM designed for the real-time and ecologically valid measurement of personal music listening. Multilevel structural equation modelling was utilised to determine predictors of emotional outcomes on both experience and listener levels. Results revealed the strongest predictors of affective change to be initial emotional state and selected music, with certain interactions of these variables critical in determining hedonic benefit or detriment. Furthermore, it was demonstrated that emotional outcomes of listening are produced almost entirely within contexts, with relatively little influence from the listener level. This comprehensive model has provided unprecedented insight into personal music listening, and the variables that are influential in producing desired emotional outcomes.

[1E] Acoustics and Psychoacoustics

LOS ANGELES 14:00-16:00 Monday 4 Aug 2014

PERCEPTUAL ROLES OF POWER FLUCTUATION FACTORS IN SPEECH

Takuya Kishida[1], Yoshitaka Nakajima[2], Kazuo Ueda[2], Gerard B. Remijn[2]
[1]Graduate School of Design, Kyushu University, Japan, [2]Dept. of Human Science/research Center for Applied Perceptual Science, Kyushu University, Japan

Time: 14:00

Factor analyses were performed by Ueda et al. [(2010). Fechner Day 2010, Padua] for critical-band filtered power fluctuations of spoken sentences in eight different languages. Three factors dividing speech into four frequency bands appeared consistently across all these languages. However, the perceptual role of each factor had not yet been clarified. We thus conducted a listening experiment, in which listeners indicated what they heard in sentences containing only information from obtained power-fluctuation factors. The number of factors was changed from one to nine. We developed a new method of factor analysis for this experiment, which we call zero-shifted factor analysis. The origin of eigenvectors was shifted from the gravity center to the acoustic silent point, i.e., the zero point. We also introduced a cepstral analysis to extract a smooth temporal change of spectral envelopes. Rather than three factors, zero-shifted factor analysis enabled us to obtain four stable factors from speech sounds in British English, Mandarin, and Japanese. Zero shifted factor analysis also enabled us to synthesize noise-vocoded Japanese speech from the obtained factors. We performed a listening experiment to determine the number of factors necessary for reasonably intelligible speech perception. Participants answered what they heard avoiding guessing, after three repetitions of each stimulus. Mora identifications were compared among the nine factor-number conditions. The mora identification performances improved as the number of factors increased. When stimuli consisted of one or two factors, few morae in the sentences could be identified by the listeners. When the stimuli consisted of three factors, by contrast, 68% of the morae were identified. When four to nine factors were included in the stimuli, mora identifications were over 85%. Four factors, which accounted for most of the power-fluctuations of speech sounds, were obtained in common from British English, Japanese, and Mandarin speech. The present experiment showed that these four common factors play important roles in speech perception.

THE INFLUENCE OF PERFORMER ROLES ON TIMBRE BLENDING BETWEEN BASSOON AND HORN PLAYERS

Sven-Amin Lembke[1], Scott Levine[1], Martha De Francisco[1], Stephen Mcadams[1]
[1]Centre for Interdisciplinary Research on Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Canada

Time: 14:30

1. Background: In orchestral music, instruments are commonly paired to achieve a blended timbre. Psychoacoustical research on wind instruments shows that their most prominent spectral traits, termed 'main formants', are important to the perception of blend. In an investigation of parametric variations of formant location, higher blend was achieved when an instrument remained at or below the formant frequency of a constant reference instrument. In music, such a reference instrument could concern a 'leading' performer that other performers follow.

2. Aims: We investigate the timbral adjustments two musicians employ in achieving blend during performance and whether assigned roles between performers establish a reference toward which timbre adjustments are oriented.

3. Method: In a behavioral experiment, pairs of bassoon and horn players performed musical excerpts together with the aim of achieving blend. Among other independent variables, performers were assigned roles as either 'leader' or 'follower', based on a within-participants design. Time analysis of timbre adjustments takes measures of spectral features (formant location, spectral centroid) and potential covariates (pitch, dynamics) into account. Each measure forms three dependent variables, quantifying the averaged magnitude, the temporal variability, and coordination between performers.

4. Results: Musicians exhibit lower main formants and spectral centroids performing as followers than as leaders. These adjustments toward 'darker' timbres fall slightly below the leaders' spectral features. Furthermore, spectral measures covary with pitch and dynamics, with the pitch variation across musical excerpts evoking spectral changes that are beyond the performers' control. On the other hand, the consistent tendency for 'darker' timbres to occur with 'softer' dynamics suggests that performers use slight shifts in dynamics to achieve the desired change in timbre.

5. Conclusions: Formant relationships between instruments are shown to be relevant to musical performance of blended timbres: performers acting as leaders function as the reference based on which followers orient their timbre to be 'darker'. In musical practice, timbre adjustments have inherent links to variations in pitch and dynamics, with the latter being used by performers to control timbre.

A QUANTITATIVE ANALYSIS OF AUDITORY STIMULI: NEW RESULTS AND INSIGHTS

Jessica Gillard[1], Michael Schutz[1]
[1]Mcmaster Institute for Music and the Mind, Canada
Time: 15:00

1. Background: The sounds used in auditory perception research typically employ "flat"' temporal structures exhibiting abrupt onsets and offsets. However, recent studies raise intriguing questions about whether these sounds are processed differently than sounds with more natural temporal characteristics (Grassi & Casco, 2009; Schutz, 2009; Vallet, Shore, & Schutz, under review).

2. Aims: This empirical study explores the degree to which auditory research employs sounds lacking real-world characteristics properties, a concern noted previously by both neuroscientists (Joris, Schreiner, & Rees, 2004) and psychologists (Gaver, 1993). Our goal is to explore whether models and theories may be built on sounds failing to trigger the perceptual processes used in everyday listening.

3. Method: Our survey classified the temporal structures of sounds from three journals: Music Perception, Attention, Perception & Psychophysics and Hearing Research. We separately analyzed each of the 587 experiments within these 310 articles, classifying their stimuli into one of 5 categories: flat (abrupt onset/offset); percussive (natural impact sounds); click/click train; undefined (i.e. lacking definition of temporal structure); and other (any other defined temporal structure).

4. Results: Curiously, "undefined" sounds were the most prominent, used in approximately 42% of these 587 experiments. This lack of definition was not journal-specific, with 35%, 54%, and 35% of undefined sounds in these three journals, respectively. This did not indicate a lack methodological detail, as 83% of sounds with undefined temporal structures employed defined spectral structures, and 68% were delivered over defined models of headphones/speakers. Less than 16% of sounds exhibited natural temporal structures (of the defined sounds, 32% were flat, 7.5% clicks/click trains, 7.4% other, and 12% percussive).

5. Conclusions: The surprising outcome that temporal structure has not merited definition in auditory perception experiments in prominent journals raises intriguing questions. We suspect most sounds with undefined temporal structures were in fact flat. However, if so then 73% of sounds encountered lacked key features of real-world sounds (i.e. meaningful offset information). This raises numerous questions given that patterns of influence (Schutz, 2009) and previously documented illusions (Grassi & Casco, 2009) change markedly when assessed with tones exhibiting "flat" vs. "natural" profiles such as those found in impact sounds.

PSYCHOACOUSTICAL AND EMOTIONAL CUES IN GERMAN POPULAR MUSIC

Michael Oehler[1], Christoph Reuter[2], Isabella Czedik-Eysenberg[2], Niklas Hill[1], Michael Ziethen[1]
[1]University for Media and Communication Cologne (MHMK), Germany, [2]University of Vienna, Musicological Institute, Austria
Time: 15:30

1. Background: In a recent study Schellenberg and Schewe (2012) analyzed musical characteristics in American popular music that are often related to emotions music conveys/induces, especially mode and tempo (Hunter & Schellenberg, 2010). As the authors focused on a long-term perspective, Billboard magazine's year-end charts (1965-2009) were used as sample (Pettijohn & Sacco, 2009). A significant increase in the use of minor mode and a decrease in average tempo over the years was found. Schellenberg and Schewe conclude that music has become more sad sounding and emotionally ambiguous.

2. Aims: The aim of our study was (a) to examine, if the effects found are also evident if a different set of popular music recordings is used (German year-end charts). We further wanted to know if (b) psychoacoustical parameters as dynamic range, spectral centroid, roughness or inharmonicity are similar or even better cues to emotion in music (Coutinho & Dibben, 2013; Nagel et al., 2008) and are therefore better predictors of the recording year than mode and tempo.

3. Method: The German year-end charts (Media Control/Musikmarkt) were used as sample (1,120 recordings from 1965-2012). Like in the original study, mode and tempo were documented by musically trained experts. The psychoacoustical parameters were extracted with MIRtoolbox (Lartillot & Toiviainen, 2007).

4. Results: There was a positive association between year of recording and mode (r_{pb}=.294, p<.0001), but no significant correlation between year of recording and tempo (r=.05, p=.096) as in the original study. A multiple regression analysis showed an increase in the use of minor mode over the years (25.1% variance explained; p<.0001), whereas tempo did not decrease significantly (p=.163). However dynamic range (41.5% variance explained, p<.0001), roughness (36.2% variance explained, p<.0001) and spectral centroid (19% variance explained, p<.0001) made a significant unique contribution to the regression model.

5. Conclusions: The results support the findings of Schellenberg and Schewe that minor mode increased significantly over the years in popular music. Additionally some psychoacoustic features have proven to be good predictors of the recording year and support the hypothesis that there is an increase in recordings with ambiguous emotional status from 1965 until today.

Session 2

[2A] Music and Timbre

SEOUL 16:30-18:30 Monday 4 Aug 2014

TIMBRAL FEATURES CONTRIBUTING TO PERCEIVED AUDITORY AND MUSICAL TENSION

Morwaread Farbood[1], *Khen Price*[1]
[1]*New York University, USA*
Time: 16:30

ABSTRACT

This paper presents two experiments exploring the contribution of timbral features to tension perception. The features examined were inharmonicity, roughness, spectral centroid, spectral deviation, and spectral flatness. The goal of the experiments was to systematically examine how changes in these features contribute directly to changes in perceived tension. In Experiment 1, the stimuli gradually morphed from two extreme states, e.g., "low" inharmonity to "high" inharmonicity. Participants were asked to judge how they felt tension was changing over the course of the sounds. In Experiment 2, the task was to compare static renderings of the low and high states of each feature separated by a silence interval. Listeners were asked to select which of the two sounds was perceived to be more tense. The results of Experiment 1 were complex and inconsistent; the process of morphing introduced certain artifacts that appeared to have influenced listener perception. Furthermore, the process of morphing made it impossible to control for all other features while examining one particular feature. The results of Experiment 2, on the other hand, were very clear—responses clearly indicated that perceived increases in tension correlated to increases in all five features.

1. INTRODUCTION

Timbre is an auditory feature that has received relatively little attention in empirical work examining musical tension. This is in part due to the general difficultly in defining the perceptual dimensions of timbre. While contributions of features such as harmony, loudness, melodic expectation, pitch height, and tempo have been examined thoroughly in past studies, it is still unclear how (and which) timbral features contribute most directly to musical tension. The timbral features that have been examined in past work (from widely varying methodological approaches) include roughness, brightness, spectral flatness, and density (Dean & Bailes, 2010; Helmholtz, 1877; Hutchinson & Knopoff, 1978; Krumhansl, 1996; Nielsen 1987; Plomp & Levelt, 1965; Pressnitzer, McAdams, Winsberg, & Fineberg, 2000).

This study presents two experiments that examine the contribution of five specific timbral features on auditory and musical tension perception: inharmonicity, roughness, spectral centroid, spectral deviation, and spectral flatness. Inharmonicity is a feature that is based on how partials are offset from integer multiples of the fundamental frequency of a pitch. Roughness, described by Plomp and Levelt (1965) as sensory dissonance roughness, is present when pairs of sinusoids are close enough in frequency such that the listener experiences a beating sensation. Spectral centroid is often associated with the perceived brightness of a sound (although the two are not synonymous), and is defined as the geometrical mean of the energy found in the different frequency bins that are produced by a fast Fourier transform (FFT) or any other applicable transformation between the time and frequency domains. Spectral deviation (also termed spectral spread) is attained by calculating the spread of the energy distribution across the spectrum. For example, a pure tone will have no spectral deviation. The spectral flatness (noisiness) of a signal corresponds to how similar its spectrum is to white noise. Low spectral flatness corresponds to more pitched sounds.

The goal of this study was to systematically examine how changes in these features contribute directly to changes in perceived tension. The "directionality" of features was also examined; that is, whether increases in the features contribute to increases in tension, or if this relationship is not clear.

2. GENERAL METHOD

For both experiments, two states were generated for each feature: one with a minimum of that particular character (state A) one with a maximum of that particular characteristic (state B). Since loudness is known to be a dominant factor in auditory perception experiments, loudness was equalized for all stimuli. When possible, an effort was made to avoid covariance between features when synthesizing the stimuli, with special attention given to keeping spectral centroid constant (in the cases where it was not the targeted feature). This is due to the established importance of this feature in timbre discrimination experiments.

Overall intensity level equalization was automatically done using the Echonest API (Jehan, 2010) so that loudness differences across stimuli were less than 1 dB. The Genesis Loudness Toolbox (Genesis, 2009) for Matlab was used to obtain time-dependent loudness measurements for each synthesized timbre for verification purposes. Amplitude envelopes and changes in intensity were applied to compensate for temporal changes in loudness, as well as differences in loudness between stimuli. The timbral features were analyzed using the MIRtoolbox (Lartillot, Toiviainen, & Eerola, 2008). Filters and noise used in the generation of spectral flatness stimuli were created using the Matlab DSP toolbox.

3. EXPERIMENT 1

3.1. Method

For each feature, a stimulus was created that morphed from state A to B, B to A, or did not change over time. In addition there were four different durations for each stimulus: 1, 5, 10, and 15 secs, resulting in a total of 60 stimuli (5 features x 3 change types x 4 durations). Fifty subjects, mostly musicians, took part in the study. The stimuli were presented to listeners over headphones in random order. Participants were asked to judge how they felt tension changed over the course of each sound by selecting one of the following multiple choice responses: increasing, decreasing, no change, or none of the above.

3.2. Results

The results of Experiment 1 were inconsistent and contradictory. There was few clear trends or "directionality" indicated in the responses. It appeared that the process of morphing introduced certain artifacts such as unintended pitch glides that appeared to influenced listener perception; this was particularly a problem for inharmonicity. Increase in inharmonicity was most strongly correlated with increase in tension when the stimuli duration was longer, most likely because the pitch shifts became slower and less noticeable.

Both pitch and loudness appeared to be an issue for spectral flatness due to the perception of crescendos and decrescendos resulting from the source separation of noise and pitch (perceived fading or increase in one or the other). Depending on which aspect (noise vs. tone) the listener's attention was drawn to, either an increase or decrease in loudness could have been salient. This would explain the conflicting results.

A perceived change in loudness and/or pitch might have also affected the results for spectral centroid, although from a qualitative perspective, pitch was not distinctly tonal in the stimuli. Spectral centroid had the most scattered and difficult to interpret results.

In the case of stimuli of longer durations, the unchanging versions were not necessarily static in loudness from a psychoacoustic perspective—having a constant sound does not necessarily translate to static perceived loudness, especially if the sound is unpleasant and tense to begin with. This would partially explain the results for spectral deviation, where tension increases were frequently indicated by listeners despite the fact that the sound remained unchanged.

Although spectral centroid was held relatively constant for stimuli focusing on the other four features, those other features were not completely controlled for. This might have further complicated the results. See Figures 1-5 (left column) for graphs showing the response profiles. Given these mixed results, a second experiment was designed to try to avoid the pitfalls encountered in morphing sound from one state to another.

4. EXPERIMENT 2

4.1. Method

Experiment 2 featured stimuli designed to compare static versions of states A and B separated by 1.2 sec of silence. In addition, there were three different pitch registers utilized for each condition (low, mid, high). For all A/B state pairs there were four possible orderings: AB, BA, AA, and BB.

Forty-six subjects, mostly musicians, took part in the study. The stimuli were presented to listeners in random order over headphones. Participants were asked to judge which of the two sounds (first or second) was more tense or if they sounded the same.

4.2. Results

Unlike the case for Experiment 1, the results of Experiment 2 were very consistent. Increases in all the features corresponded strongly with changes in tension (Figures 1-5, right column). The only interaction effect with pitch register occurred in the case of spectral centroid. Those results indicated that the higher the pitch register, the stronger the correlation between increases in spectral centroid and tension.

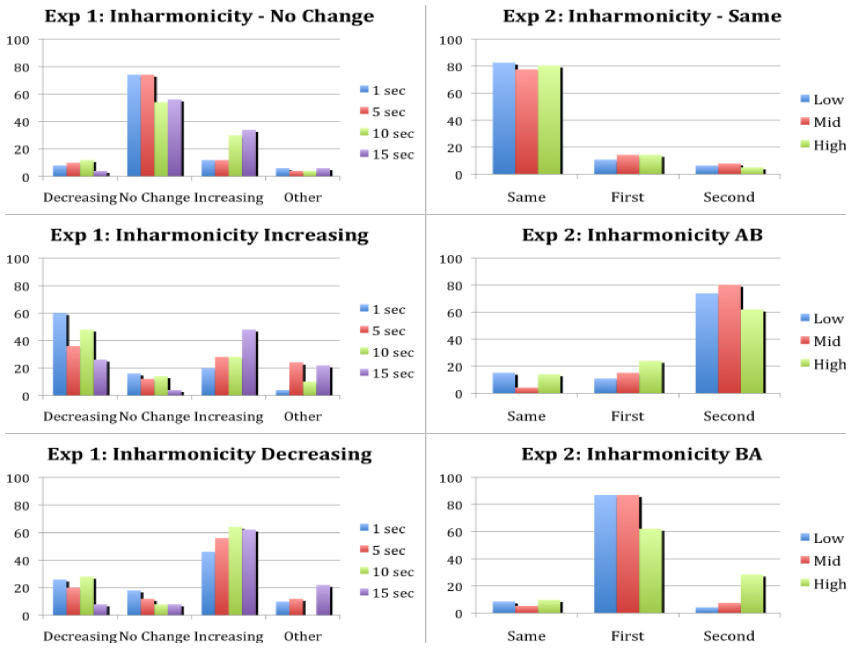


Figure 1: Response profiles for stimuli featuring changing inharmonicity; graphs in the left column show results from Experiment 1, graphs in the right column from Experiment 2.

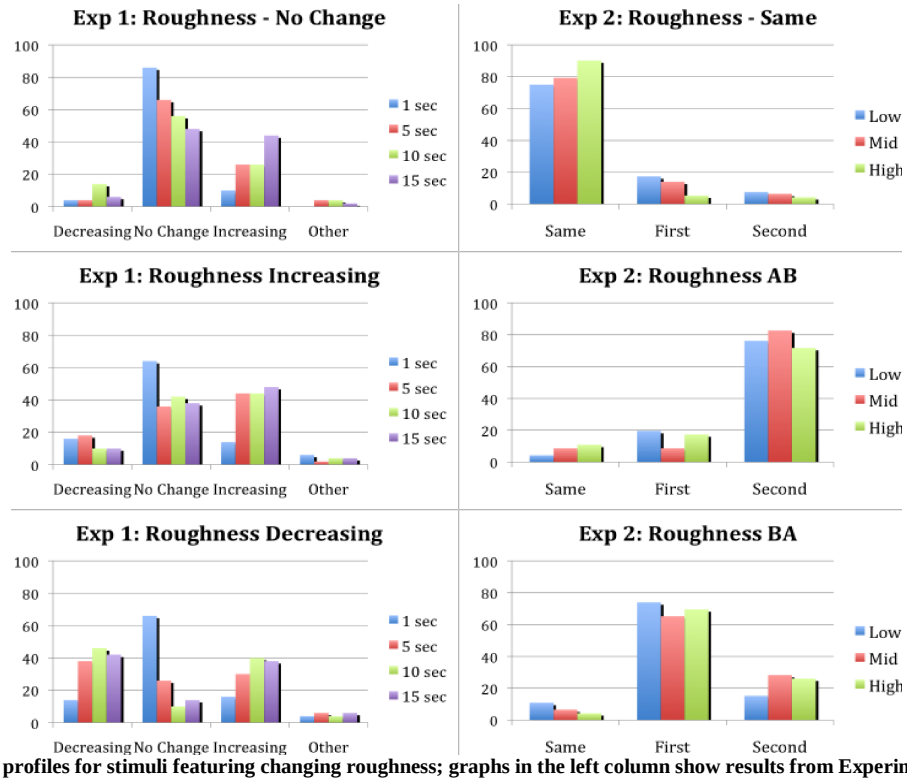


Figure 2: Response profiles for stimuli featuring changing roughness; graphs in the left column show results from Experiment 1, graphs in the right column from Experiment 2.

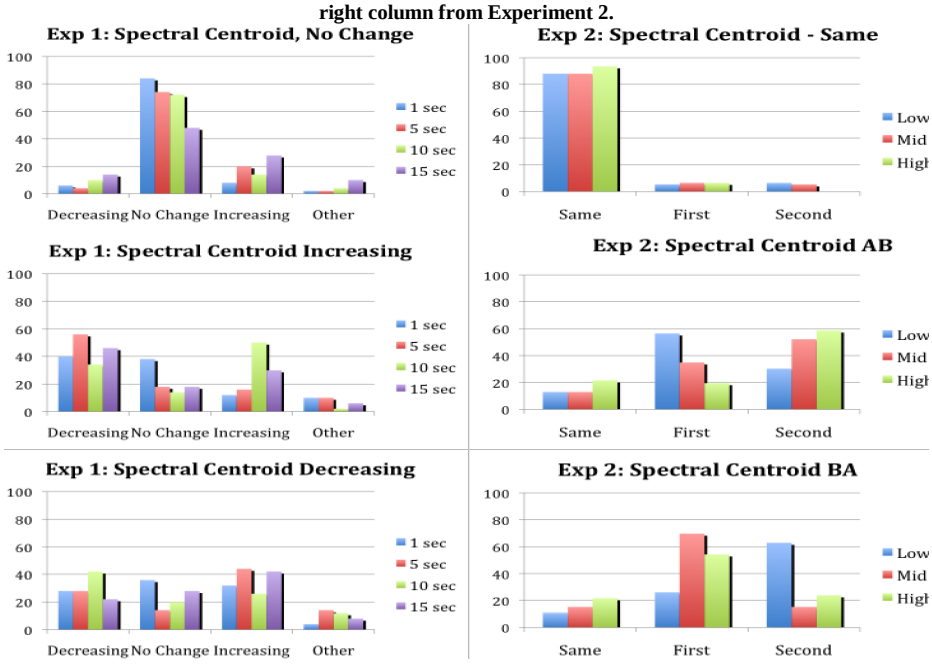


Figure 3: Response profiles for stimuli featuring changing spectral centroid; graphs in the left column show results from Experiment 1, graphs in the right column from Experiment 2.

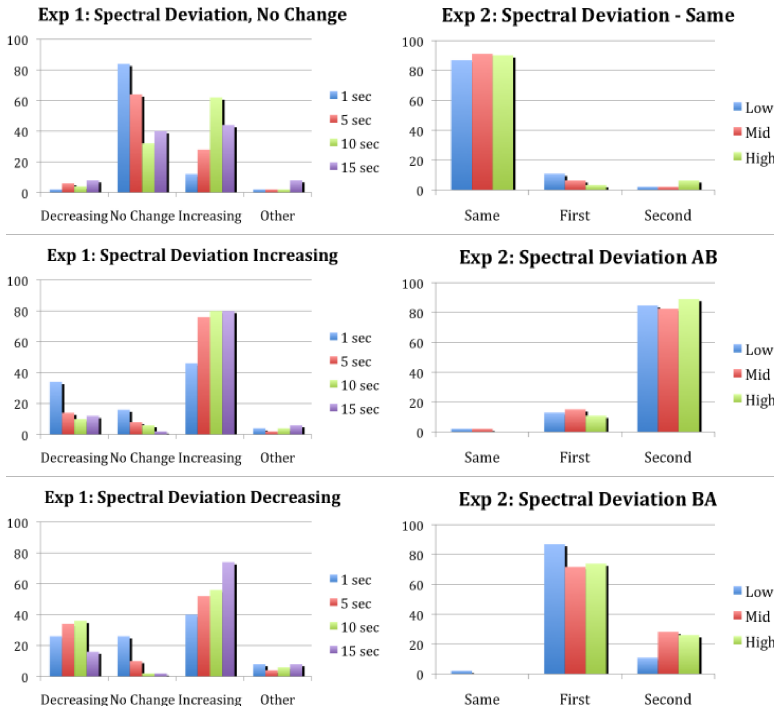


Figure 4: Response profiles for stimuli featuring changing spectral deviation; graphs in the left column show results from Experiment 1, graphs in the right column from Experiment 2.

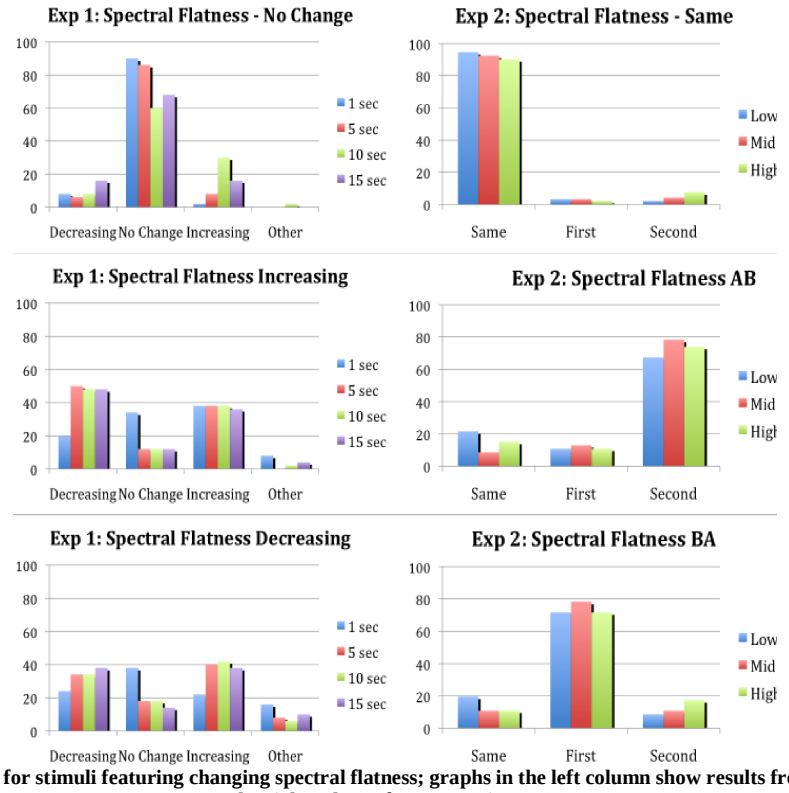


Figure 5: Response profiles for stimuli featuring changing spectral flatness; graphs in the left column show results from Experiment 1, graphs in the right column from Experiment 2.

5. CONCLUSIONS

The five timbral features examined in this study—inharmonicity, roughness, spectral centroid, spectral deviation, and spectral flatness—were shown to contribute to perceived tension. However, changes in some of these parameters were difficult to isolate because of feature covariance. As such, these conclusions were only apparent in the results of Experiment 2. The results of Experiment 1 were complex and inconsistent due to the inability to control for all features as well as pitch and loudness. Experiment 2 used static versions of the two extreme states for each feature, resulting in much clearer response profiles. In all cases, an increase in each feature corresponded to an increase in perceived tension.

The next step is to explore precisely how these features covary in order to model how dynamic timbral changes influence tension perception. Additional experiments using more complex stimuli—particularly musical stimuli where other musical features influencing tension such as harmony and melodic contour are involved—are the next directions to explore in future work.

6. REFERENCES

Dean, R. T., & Bailes, F. (2010). Time series analysis as a method to examine acoustical influences on real-time perception of music. Empirical Musicology Review, 5*(4), 152–175.*

Gensis (2009). Loudness Toolbox. Retrieved from http://www.genesis-acoustics.com/en/index.php?page=32

Helmholtz, von, H. L. F. (1877). Die Lehre von der Tonempfindungen als physiologische Grundlage für die Theorie der Musik. New York: Dover.

Jehan, T. (2010). Echo Nest Remix API. Retrieved from http://developer.echonest.com/client_libraries.html

Krumhansl, C. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. Music Perception, 13*(3), 401–432.*

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for music information retrieval. In Studies in Classification, Data Analysis, and Knowledge Organization: Data Analysis, Machine Learning and Applications *(pp. 261–268). Springer.*

Nielsen, F. (1987). Musical "tension" and related concepts. The Semiotic Web '86: an International Yearbook, *491–514.*

Plomp, R., & Levelt, W. J. M. (1965). Tonal Consonance and Critical Bandwidth. Journal of the Acoustical Society of America, 38, *548–560.*

Pressnitzer, D., McAdams, S., Winsberg, S., & Fineberg, J. (2000). Perception of music tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. Perception & Psychophysics, 62*(1), 66–80.*

CONTEXT EFFECTS IN THE COGNITIVE SEQUENCING OF MUSICAL TIMBRE

Kai Siedenburg[1], Stephen Mcadams[1]
[1]Centre for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University, Montreal, QC, Canada
Time: 17:00

1. Background: Musical timbre denotes the set of auditory attributes that lend sounds their particular qualia and enable the auditory recognition of musical instruments. There are a variety of musical style-systems in which sequential timbral structures play important parts. Nonetheless, little is known about the mechanisms that determine the cognition of timbre in such musical contexts.

2. Aims: The current study systematically investigated the impact of contexts on short-term recognition of timbral sequences. Here, context is understood in a two-fold manner, relating to serial order ("where"), and to the similarity relations between items in the sequence ("what").

3. Method: In each trial, subjects (N=22) listened to a standard sequence and a comparison sequence delayed by 3 s, both containing the same four distinct timbres. They were asked to judge whether standard and comparison were of same or different order. Within-subject factors were the serial position of swaps of items (1&2, 2&3, 3&4, 2&4) and similarity between the timbres that swapped (low, medium, high). Timbres and their mean pairwise dissimilarity judgments stemmed from McAdams et al. (1995, Psychol. Res.).

4. Results: A repeated-measures analysis of variance of d' scores demonstrated a significant effect of position ($p = .022$, partial eta squared =.140) due to a recency effect (but with no significant primacy effect); similarity generated a strong effect ($p < .001$, partial eta squared =.608) with low similarity yielding significantly higher recognition scores compared to medium and high. Most crucially, however, 94% of the variance in z-transformed proportions of listeners responding "same" (averaged over serial positions) was predicted by an additive combination of similarity of swap and timbral heterogeneity (mean pairwise distance between all items in a sequence) in a multiple linear regression analysis.

5. Conclusions: The current study highlights the importance of two different aspects of context – "the where and the what" - in short-term memorization and recognition of timbre. The results suggest that the cognition of timbre sequences, sonic gestalts extending over time, can be efficiently modeled using the parameter's topology of similarity as a starting point.

CONGRUENCY OF COLOUR ASSOCIATIONS TO URBAN SOUNDSCAPES, RESTAURANTS, AND ELECTROACOUSTIC SOUNDSCAPE COMPOSITIONS.

Permagnus Lindborg[1,2]
[1]Nanyang Technological University, Singapore, [2]KTH Royal Institute of Technology, Sweden
Time: 17:30

1. Background: Soundscapes are the auditory parts of multimodal environments (physical or virtual). We propose that while they may have differing degree of design intent: from non-designed (e.g. pristine, natural, urban areas), via semi-designed (servicescapes, restaurants), to fully designed (film, games, concert music), people approach them with similar listening strategies (Tuuri & Eerola 2012). The ecological model of perception explains crossmodal interactions as biological adaptations. A sound from an identified source may elicit a predictable affective response, but an acoustically equivalent but unidentifiable sound might not (Västfjäll 2013). A richer understanding of soundscape perception could be gained by considering visual associations (e.g. Haverkamp 2009, Giannakis & Smith 2001) of different kinds, in addition to semantic scales (e.g. Axelsson et al. 2010).

2. Aims: The aim with the present study was to investigate a) correlations between psychoacoustic metrics, and visual and semantic responses; and b) congruency of such multimodal responses across soundscapes of varying degree of design intent.

3. Method: In separate perceptual experiments, we used binaural reproductions of urban soundscapes (N=43), restaurants (N=31), and electroacoustic soundscape compositions (N=20, underways). Respondents made colour associations using a CIELAB model, provided free-form single-word tagging of perceived sound sources, and spatially sorted stimuli using a graphical interface. The semantic associations, classified by 5 independent raters, yielded a Valence-Arousal estimate for each stimulus. The dissimilarity between stimuli was analysed with isometric multidimensional scaling, and compared with psychoacoustic metrics computed from recordings (loudness, roughness, sharpness, fluctuation strength).

4. Results: Results from the first two experiments show that louder soundscapes were associated with more reddish and darker colours. At the same time, loudness (and to a lesser extent sharpness) correlated with Arousal and Valence estimates, as expected, but also with the way stimuli were sorted in the graphical (spatial sorting) interface.

5. Conclusion: Responses to soundscapes in these experiments were consistent across visual (colour and spatial sorting) and semantic modalities. Further results, including the third experiment about electroacoustic soundscape compositions, will be presented at the conference.

[2B] Listening as a Creative Process

KYOTO 16:30-18:30 Monday 4 Aug 2014

HARMONY PERCEPTION BY PERIODICITY DETECTION

Frieder Stolzenburg[1]
[1]*Harz University of Applied Sciences, Germany*
Time: 16:30

1. INTRODUCTION

Numerous approaches tackle the question, how musical consonance/dissonance perception can be explained. Many corresponding empirical studies reveal a clear preference ordering on the perceived consonance/dissonance of common triads in Western music, e.g. major < minor (Roberts, 1986). Early mathematical models expressed musical intervals, i.e. the distance between two pitches, by simple fractions. Newer explanations base on the notion of dissonance, roughness, instability or tension (Helmholtz, 1862; Hutchinson and Knopoff, 1978, 1979; Cook and Fujisawa, 2006). They correlate better to empirical results on harmony perception, but still can be improved.

A general theory of harmony perception should be applicable to musical harmonies in a broad sense, i.e. to chords and scales. It should consider results from neuroacoustics and psychophysics on auditory processing, e.g. that periodicities of complex chords can be detected in the brain (Langner, 1997; Lee et al., 2009). Furthermore, the correlation between the predicted and the perceived consonance/dissonance from empirical studies should be the highest possible. Several empirical experiments on harmony perception have been conducted, where participants are asked to listen to and rate musical harmonies. The corresponding results (Malmberg, 1918; Johnson-Laird et al., 2012; Temperley and Tan, 2013) can be used to evaluate theories on harmony perception.

2. MAIN CONTRIBUTION

We apply recent results from psychophysics and neuroacoustics consistently, obtaining a fully computational theory of consonance/dissonance perception. We focus on periodicity detection including autocorrelation analysis, which can be identified as a fundamental mechanism to music perception, and exploit in addition that the just noticeable difference of human pitch perception is about 1% for the musically important low frequency range (Zwicker et al., 1957). We transfer the concept of periodicity pitch to chords and scales by considering relative periodicity, i.e. the approximated ratio of the period length of the chord relative to the period length of its lowest tone component. The hypothesis is that the perceived consonance of a musical harmony decreases with the relative (logarithmic) periodicity.

In this context, we adopt the equal temperament as reference system for tunings here. The frequency ratios in equal temperament are irrational numbers (except for the unison and its octaves), but for computing periodicity they must be fractions. We thus consider tunings with rational frequency ratios. The oldest tuning with this property is probably the Pythagorean tuning which, however, is not in line with the results of psychophysics, in particular, that the just noticeable difference of pitch perception is about 1%. We therefore use the rational tuning, which takes the fractions with smallest possible denominators, such that the relative deviation with respect to equal temperament is just below a given percentage d. Its frequency ratios deviate only slightly from just tuning, namely only for the tritone

and the minor seventh. Just tuning can be understood as rational tuning with only slightly greater maximal relative deviation d=1.1%. We adopt it as well as the rational tuning as underlying tuning in our analyses. Table 1 shows the respective frequency ratios.

Table 1: Table of frequency ratios for different tunings. Here, k denotes the number of the semitone corresponding to the given interval. In parentheses, the relative deviation of the respective frequency ratio from equal temperament is shown. The maximal deviation d for the rational tuning listed here is 1%.

interval	k	rational tuning	just tuning
unison	0	1/1 (0,00%)	1/1 (0,00%)
minor second	1	16/15 (0.68%)	16/15 (0.68%)
major second	2	9/8 (0.23%)	9/8 (0.23%)
minor third	3	6/5 (0.91%)	6/5 (0.91%)
major third	4	5/4 (-0.79%)	5/4 (-0.79%)
perfect fourth	5	4/3 (-0.11%)	4/3 (-0.11%)
tritone	6	17/12 (0.17)	7/5 (-1.01%)
perfect fifth	7	3/2 (0.11%)	3/2 (0.11%)
minor sixth	8	8/5 (0.79%)	8/5 (0.79%)
major sixth	9	5/3 (-0.9%)	5/3 (-0.9%)
minor seventh	10	16/9 (-0.23%)	9/5 (1.02%)
major seventh	11	15/8 (-0.68%)	15/8 (-0.68%)
octave	12	2/1 (0.00%)	2/1 (0.00%)

3. RESULTS

The predictions of the periodicity-based approach obtained for dyads, common triads, and diatonic scales all show highest correlation with empirical results (Schwartz et al., 2003; Johnson-Laird et al., 2012; Temperley and Tan, 2013), not only with respect to the ranks, but also with the ordinal values of the empirical ratings of musical consonance, in particular, when logarithmic periodicity is employed, which can be motivated by the topological organization of the periodicity coding in the brain. Interestingly, the logarithmic periodicity of the complete chromatic scale is within the biological bound of 8 octaves, which can be represented in the neuronal periodicity map in the brain.

In our analyses, we correlate the empirical and the theoretical ratings of harmonies. Since in most cases only data on the ranking of harmonies is available, we mainly correlate rankings. Nevertheless, correlating concrete numerical values yields additional interesting insights (see below). For the sake of simplicity and consistency, we always compute Pearson's correlation coefficient r, which coincides with Spearman's rank correlation coefficient on rankings, provided that there are not too many bindings, i.e. duplicate values.

Table 2 shows the perceived and computed relative consonance of dyads (intervals). The correlations of the empirical rating with the sonance factor and with relative or logarithmic periodicity show the highest correlation (r=.982). Table 3 shows the perceived and computed relative consonance of common triads. There are several empirical studies on the perception of common triads. But since the experiments conducted by Johnson-Laird et al. (2012) are the most comprehensive, because they examined all 55 possible three-note chords, we adopt this study as reference for the empirical ranking here. Nonetheless, all studies are consistent with the following preference ordering on triads: major < minor < suspended < diminished < augmented, at least for chords in root position. However, the ordinal ratings of minor and suspended chords do not differ very much. Again, the analysis reveals highest correlations for relative and logarithmic periodicity, if the underlying tuning is psychophysically motivated. Roughness (Hutchinson and Knopoff, 1978, 1979) and the sonance factor (Hofmann-Engl, 2004, 2008) yield relatively bad predictions on the perceived consonance of common triads.

The data sets in Johnson-Laird et al. (2012) suggest further investigations. So, Table 4 shows the analysis of all possible three-tone chords in root position. As one can see, the correlation between the empirical rating with the predictions of the dual-process theory is very high (r=.916). This also holds for logarithmic periodicity but not that much for relative periodicity, in particular, if the correlation between the ordinal rating and the concrete periodicity values are taken. This justifies our preference of logarithmic to relative periodicity, because the former notion is motivated more by neuroacoustical results, namely that the spatial structure of the periodicity-pitch representation in the brain is organised as a logarithmic periodicity map.

Temperley and Tan (2013) investigate the perceived consonance of diatonic scales. Table 5 lists all classical church modes, i.e. the diatonic scale and its inversions. The cognitive model on the perception of diatonic scales introduced by Temperley and Tan (2013) results in a 100% correlation with the empirical data. Although the correlation for logarithmic periodicity obviously is not that good, it shows still high correlation. Nonetheless, the major scale (Ionian) appears in the front rank of 462 possible scales with 7 out of 12 tones with respect to relative and logarithmic periodicity. In addition, in contrast to more cognitive theories on harmony perception, the periodicity-based approach introduced in this paper does not presuppose any principles of tonal music, e.g. the existence of diatonic scales or the common use of the major triad. They can be derived from underlying, more primitive mechanisms, namely periodicity detection in the human (as well as animal) brain.

4. REFERENCES

Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. Perception and Psychophysics, 58(1):125–141.

Cook, N. D. and Fujisawa, T. X. (2006). The psychophysics of harmony perception: Harmony is a three-tone phenomenon. Empirical Musicology Review, 1(2):1–21.

Gill, K. Z. and Purves, D. (2009). A biological rationale for musical scales. PLoS ONE, 4(12):e8144.

Johnson-Laird, P. N., Kang, O. E., and Leong, Y. C. (2012). On musical dissonance. Music Perception, 30(1):19–35.

Hofmann-Engl, L. J. (2004). Virtual pitch and its application to contemporary harmonic analysis. Chameleon Group Online Publication.

Hofmann-Engl, L. J. (2008). Virtual pitch and the classification of chords in minor and major keys. In Miyazaki, K., Hiraga, Y., Adachi, M., Nakajima, Y., and Tsuzaki, M., editors, Proceedings of 10th International Conference on Music Perception and Cognition, Sapporo, Japan.

Hutchinson, W. and Knopoff, L. (1978). The acoustic component of western consonance. Interface/Journal of New Music Research, 7(1):1–29.

Hutchinson, W. and Knopoff, L. (1979). The significance of the acoustic component of consonance in Western triad. Journal of Musicological Research, 3(1-2):5–22.

Langner, G. (1997). Temporal processing of pitch in the auditory system. Journal of New Music Research, 26:116–132.

Lee, K. M., Skoe, E., Kraus, N., and Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. The Journal of Neuroscience, 29(18):5832–5840.

Malmberg, C. F. (1918). The perception of consonance and dissonance. Psychological Monographs, 25(2):93–133.

Schwartz, D. A., Howe, C. Q., and Purves, D. (2003). The statistical structure of human speech sounds predicts musical universals. The Journal of Neuroscience, 23(18):7160–7168.

Temperley, D. and Tan, D. (2013). Emotional connotations of diatonic modes. Music Perception, 30(3):237–57.

Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). Critical band width in loudness summation. Journal of the Acoustical Society of America, 29(5):548–557.

interval		emp. rank	roughness	sonance factor	similarity	rel. periodicity
unison	{0, 0}	1	2 (0.0019)	1-2 (1.000)	1-2 (100.00%)	1-2 (1.0)
octave	{0, 12}	2	1 (0.0014)	1-2 (1.000)	1-2 (100.00%)	1-2 (1.0)
perfect fifth	{0, 7}	3	3 (0.0221)	3 (0.737)	3 (66.67%)	3 (2.0)
perfect fourth	{0, 5}	4	4 (0.0451)	4 (0.701)	4 (50.00%)	4-5 (3.0)
major third	{0, 4}	5	6 (0.0551)	5 (0.570)	6 (40.00%)	6 (4.0)
major sixth	{0, 9}	6	5 (0.0477)	6 (0.526)	5 (46.67%)	4-5 (3.0)
minor sixth	{0, 8}	7	7 (0.0843)	7 (0.520)	9 (30.00%)	7-8 (5.0)
minor third	{0, 3}	8	10 (0.1109)	8 (0.495)	7 (33.33%)	7-8 (5.0)
tritone	{0, 6}	9	8 (0.0930)	11 (0.327)	8 (31.43%)	9 (6.0)
minor seventh	{0, 10}	10	9 (0.0998)	9 (0.449)	10 (28.89%)	10 (7.0)
major second	{0, 2}	11	12 (0.2690)	10 (0.393)	11 (22.22%)	12 (8.5)
major seventh	{0, 11}	12	11 (0.2312)	12 (0.242)	12 (18.33%)	11 (8.0)
minor second	{0, 1}	13	13 (0.4886)	13 (0.183)	13 (12.50%)	13 (15.0)
correlation r			.967	.982	.977	.982

Table 2: Consonance rankings of dyads. The respective numbers of semitones with respect to the Western twelve-tone system are given in braces, raw values of the respective measures in parentheses. The empirical rank is the average rank according to the summary given by Schwartz et al. (2003, Figure 6). The roughness values are taken from Hutchinson and Knopoff (1978, Appendix). For computing the sonance factor (Hofmann-Engl, 2004, 2008), the *Harmony Analyzer 3.2* applet software has been used, available at http://www.chameleongroup.org.uk/software/piano.html. For these models, always $C4$ (middle C) is taken as lowest tone.

chord class		emp. rank	roughness	instability	similarity	rel. periodicity	dual proc.
major	{0, 4, 7}	1 (1.667)	3 (0.1390)	1 (0.624)	1-2 (46.67%)	2 (4.0)	2
	{0, 3, 8}	5 (2.889)	9 (0.1873)	5 (0.814)	8-9 (37.78%)	3 (5.0)	1
	{0, 5, 9}	3 (2.741)	1 (0.1190)	4 (0.780)	5-6 (45.56%)	1 (3.0)	3
minor	{0, 3, 7}	2 (2.407)	4 (0.1479)	2 (0.744)	1-2 (46.67%)	4 (10.0)	4
	{0, 4, 9}	10 (3.593)	2 (0.1254)	3 (0.756)	5-6 (45.56%)	7 (12.0)	5
	{0, 5, 8}	8 (3.481)	7 (0.1712)	6 (0.838)	8-9 (37.78%)	10 (15.0)	6
susp.	{0, 5, 7}	7 (3.148)	11 (0.2280)	8 (1.175)	3-4 (46.30%)	5 (10.7)	7
	{0, 2, 7}	6 (3.111)	13 (0.2490)	11 (1.219)	3-4 (46.30%)	9 (14.3)	9
	{0, 5, 10}	4 (2.852)	6 (0.1549)	9 (1.190)	7 (42.96%)	6 (11.0)	8
dim.	{0, 3, 6}	12 (3.889)	12 (0.2303)	12 (1.431)	13 (32.70%)	12 (17.0)	12
	{0, 3, 9}	9 (3.519)	10 (0.2024)	7 (1.114)	10-11 (37.14%)	11 (15.3)	10
	{0, 6, 9}	11 (3.667)	8 (0.1834)	10 (1.196)	10-11 (37.14%)	8 (13.3)	11
augm.	{0, 4, 8}	13 (5.259)	5 (0.1490)	13 (1.998)	12 (36.67%)	13 (20.3)	13
correlation r			.352	.698	.802	.846	.791

Table 3: Consonance rankings of common triads. The empirical rank is adopted from Johnson-Laird et al. (2012, Experiment 1), where the tones are reduced to one octave in the theoretical analysis here. The roughness values are taken from Hutchinson and Knopoff (1979, Table 1), where again $C4$ (middle C) is taken as the lowest tone. For relative periodicity and percentage similarity (Gill and Purves, 2009), the frequency ratios from just tuning are used. The dual-process theory (Johnson-Laird et al., 2012, Figure 2) as a cognitive theory only provides ranks, no numerical raw values.

CONSONANCE AND DISSONANCE PERCEPTION IN INFANTS: THEORETICAL CONSIDERATIONS AND EMPIRICAL PROPOSAL

Nicola Di Stefano[1]
[1]*University Campus Bio-medico, Rome, Italy*
Time: 17:00

1. Background: The focus of our investigation is the perception of consonance and dissonance in children. From a theoretical point of view, the nature of consonance and dissonance has been studied in musical history since its beginning. From Plato and Aristoxenus to Helmholtz, due to the role played in the construction of the harmonic language, theorists and composers believed that the definition of these terms was preliminary to any further development of musical syntax. From an empirical point of view, the ability of perceiving consonance and dissonance has been more and more widely investigated in various subject and through different experimental procedures over the past 20 years.

2. Aims: Though there is a great convergence in terms of results, proving that the ability examined in humans has a biological basis, there is a strong need to focus on the different methods adopted and on the different interpretation of data.

3. Main contribution: We consider the difference between an inner approach, such as fMRI, where the experimenter's role is limited, and a behavioral one where the experimenter plays a decisive role. We focus particularly on a behavioral approach, the "looking time" method: its origins, use, interpretation and reliability in acoustical perception literature and research.

4. Implications and Conclusions: In conclusion, we introduce a device developed in University Campus Bio-Medico. The device is a musical toy which emits different sounds at different degrees of movement. In particular, when turned left, i.e., emits consonant intervals, when right, dissonant intervals. We present the results of the test phase, which has been conducted in a nursery on infants of about one year age. Connected to a PC, the device allows us to measure and record the frequency (how many times) and the length of time the baby "plays" the toy in a position corresponding to consonance or dissonance sounds. This procedure provides us more direct, simple and raw data, not being invasive or complex – as happens by means of fMRI – and reducing the experimenter's interpretation with respect to the traditional "looking time method" or "head turn procedure".

CREATING MUSIC THAT MOVES - CREATIVE DRAMA AND MOVEMENT AS TOOLS FOR IMPROVISATION IN MUSIC

Edgar Cardoso[1]
[1]*Universidade de Aveiro, Portugal*
Time: 17:30

ABSTRACT

This paper reflects on social learning issues of a series of workshops for children in Portugal called "Creative Piano". During the workshops, children take part in several exercises of drama and movement improvisation, where they unconsciously train their physical awareness, their immediate reaction to the circumstances and their relationship with their audience. Afterwards, they are encouraged to bring ideas that came up in the previous exercises to the piano. They are invited to improvise at the piano to convey the same thoughts they previously tried to communicate through other means. By using this process of thinking across performing arts under the aim of communication to an audience, they start to develop their improvisational skills under meaningful sound objectives.

1. INTRODUCTION

Nowadays, learning through social experience is the key for development. Albert Bandura's theory of social learning (1986) has been defended and listed as the most adequate to the principles of post-modernist science. In fact, vicarious reinforcement is the motto of more recent theories on education, stating that collaborative work is the answer for a globalized and multicultural society. Andy Hargreaves believes that collaboration and collegiality allow the whole educative community to develop, enabling a change to happen (1994, p. 186). In his work, Hargreaves adds that schools need reforms on their system, in order to face the implications of the heterogeneous student mass of the present. Teachers need to be less individualist and communicate with peers, so as to regenerate knowledge and open the discussion for school initiatives. Schools need "collaborative cultures" (Hargreaves, 1994, p. 192) that give the opportunity to all the members of the community to raise their voice and make their statement in the process of knowledge acquisition.

There has been a massive worldwide discussion about the potential for interdisciplinarity in today's schools curricula. Especially in the arts, there is an urge for the development of collaborative artistic projects to create original and fresh art products. Although the experimental work continues, there is still a lack of sense of legacy, i.e. creativity is crossing borders on the edge of the moment of performance, because this is the reality of creative artists who seek for new objects; on the other hand, schools in general show little interest in the encouragement of interactive creativity, leaving this process of transcendence to an older stage of life. The philosophy of constructivism emphasizes the autonomous process of knowledge construction, leading to a personal creative mind. But isn't creativity actually the most important source of data assimilation and connection?

Creative thinking through inter-artistic connections is the main focus of "Creative Piano", a series of workshops that were born as a collaborative project. It was thought to serve as a complement for the piano lessons, where exam requirements would take all the necessary time for creative experiences in class. Groups of 10 students from different piano teachers in a music school are gathered for a week of daily sessions of creative tasks that draw inspiration from Creative Drama and Movement. On a first part, children take part in improvisation games, where they stimulate their imagination by creating stories, scenes, pictures and actions to solve a determined task.

The second part of the sessions is when they are asked to apply concepts and ideas from drama improvisation into piano improvisation. The musical outcome of the final improvisation is then a result of an immersive experimentation with mind and body.

Through empirical findings, I believe that the immersion in various forms of art is mutually profitable. All arts can be used as a complement to each other, equipping us with creative tools that can be applied to any type of artistic expression. For example, I often find inspiration in art objects when I perform on the piano, because they give me a visible reference for the musical texture that I want to produce. Some instruction in architecture and sculpture provided me with a sense of structure, very relevant for the emotional appeal when performing music, as studies suggest (Sloboda, 1991, p. 120). On this specific project, drama and movement were used as to enrich musical education, because all of the three modes of performing arts have a heavy component for improvisation and creation of a product in the moment, for an audience.

2. THE GAME OF IMAGINATION

I developed a pedagogical methodology centered on dramatic improvisation games, which are very popular in the training programs for actors. Famous dramatist Augusto Boal would always use games to activate and connect the five senses and engender an atmosphere of fun and creativity (1993). Moreover, improvisation games in group help establish a social awareness and a creative environment through the development of social skills. Because of group experimentation, students lose the fear of the risk and feel the need to cooperate in order to accomplish a challenge.

During most of the games, children must either impersonate someone, or an object, or an animal, and they need to produce a story by delivering actions. By incorporating their ideas into their bodies, they learn to associate emotion to expression, and by clarifying their body language, they develop their ability to communicate.

We can *play* an instrument, but we can also *play* a character. Both expressions stress the value of playful experiences for creative purposes, in the sense that we can only achieve something new if we keep on trying and failing. Playing is also a representation, a fake incorporation in order to appeal another person's aesthetic perception. When it comes to dramatic games, Leenhardt classifies them as "means or oral and corporal expression so careless and yet so essential to adult life" (Leenhardt, 1974, p. 27). According to Melo (2005), they can offer an understanding of social matters, moving the participants to analyze the past and search for answers for the problems of the present.

One of the most important aspects of childhood is the playful relationship with reality, where empirical learning is naturally preferred. Drama games are appealing to children because they dive in a different world, using their imagination to represent something out of their actions (Libras, 2012, p. 27). These games appeal to the exploratory action with interaction, mixing the terms of actor and spectator. The creation of a secondary reality will help to reflect on the primary reality and the modes of behavior in real life (Cavadas, 2011, p. 14). Indeed, in a school context, the child gets motivated with challenges to her/his consciousness, bringing her/him confidence to deal with similar aspects in real life situations (Núñez & Navarro, 2007, p. 232). Incorporating symbolic codes in communication and artistic expression, children build their socio-cultural repertoire with "modes of signification of the world and of the intentional action" (Sarmento *apud* Gonçalves & Trindade, 2008, p. 3).

3. CREATIVE MINDS IN CREATIVE BODIES

Psychophysiologists have been studying for decades the interface between mind and body, assuming that "behavioral, cognitive, emotional, and social events all are mirrored in physiological processes" (Hugdahl, 1995, p. 3). Turpin adds that people internalize knowledge through psychological *processes* (mind) and physiological *correlates* (body) (1989, p. 219).

The embodiment of artistic conceptions helps the comprehension of one's own sense of expressivity. Creative drama and creative movement comprise simple strategies to develop children's subjectivity and space awareness, because of enabling mind processes to be correlated in action.

Carlos Abril (2011) reviewed the associations of music with movement in the past. Dalcroze, Kodály, Orff and Gordon were all music theorists who stressed the importance of the use of kinesthetic intelligence in music education, believing that a visual stimulus and a sense of space can help to understand the notions of rhythm, pulse, and intervals. The popular term *audiation*, created by Edwin Gordon, has to do with a body perception of sound, without which it is impossible to internalize musical rhythms and different pitches. Laban and Weikart even produced methods of movement vocabulary to apply in music learning.

Other studies (Parncutt, 2003; Costa, 2009) help proving that the phenomenon of integration of several artistic means of communication in a pedagogical context is beneficial for the instigation of creative thinking and group awareness. Body consciousness is a basic element on Creative Drama, because "it drives one to feel herself/himself in a natural space, re-finding the malleability lost by social imposition" (Cunha *apud* Dias, 2010, p. 56). Gonçalves & Trindade adhere that our body reveals our personal characteristics in all its manifestations, either postures, gestures or reactions, functioning as an "exhibition of our singularity as individuals" (2008, p. 2). In this context, I conclude the potential of the combination of music, drama and movement, because they all together help us find our identity and express it.

4. IMPROVISING STORIES

My pedagogical approach to improvisation is a subjective one, or even inter-subjective. Improvised performances in these workshops start with an artistic income provided by a non-musical reference. It can be a concept, an action, an image, a story or a scene. Experimentation on the piano is primarily inspired by an individual perspective, but after it is expressed to the group of peers, it will generate new models for others to question about. It represents a collective experience carried out of individual expressions.

Stories constitute the guidelines for dramatic experimentation, because they offer a plot and suggest actions. Narration is also present in music, with the association of imagery to musical works, or even more objectively in relation to programmatic music. It can be used with stylistic and cultural fundaments to convey a personal interpretation of a musical work in a specific historical context. In both cases, narration organizes the artistic thinking, giving it an aesthetic coherence, with enough creative freedom.

Stories can be creatively built or evoked by empiric memory. Studies (Gersie, 1993; Pendzik, 2006) reveal how the revival of life memories can originate new ideas and new perspectives on the present. That is the process of acting improvisation, where children make use of the values that they empirically acquired to face situations in the moment. The post-action reflection is the moment of skill development, when they decide if their mode of behavior brought good or bad consequences. Their unawareness of the learning process keeps them from deflecting. Stories have a potential to be used as codes of signification in any type of artistic improvisation. There has been experimentation (Levinson, 2004; Bernhardt, 2006) with stories on piano improvisation, where the musical material is coded into the story in different forms, enabling children to *play* around with the sound through narrative imagination. Other authors (Hyry-Beihammer, 2011; Bruner *apud ibidem*) stress the emergency of a constructivist method of piano teaching, where dialogue and discussion with the student build argumentative skills on aesthetic perception. The narrative approach is extremely useful for the cultural appreciation and recognition, as Bruner declares: "I understand that our culture – the everyday life around us – affects our manner of thinking, and that we learn culture by telling stories and listening to stories within culture" (Bruner *apud* Hyry-Beihammer, 2011, p. 207).

One of the most common criticisms of subjective improvisation is the lack of musical tools for ideas to develop. Indeed, children always ask before improvising: "How can I portrait that into music?" But the answer is really what this creative process is all about. It is about generating musical textures that might resemble an individual idea. The richness of the artistic creativity is in the plural meanings and diverse interpretations to one another, according to their empirical context. And yet, "the principles of true art is not to portray, but to evoke" (Kosinski, 1992). Through the challenge of transposing ideas into sounds, children create their own language for improvisation, producing a truly individual work, regardless of aesthetic preconceptions.

5. FINAL THOUGHTS

The concepts of inter-subjectivity, interaction and interdisciplinarity are the key aspects of the contemporary ideologies on an ideal system of education. The paradigm of social constructivism is being widely discussed to bring a collaborative culture to schools that need a refresh in their methods of teaching. In a world with permanent knowledge review and renovation, teachers must prioritize learning instead of teaching, by guaranteeing that students are equipped with tools to build autonomously their knowledge. Collaborative work in schools shall not be exclusively among teachers, but also among students. Vygotsky stressed the importance of social learning prior to individual development (Panofsky, 2003). Today it is essential that children are brought together in group to the learning process, in order to respect, accept and understand each other and to argue, discuss and interpret in a safe democratic environment.

The use of narration in artistic improvisation may facilitate the comprehension of culture and style, developing also social criticism and imagery awareness with the empiric process of performing. Different artistic resources must be applied in improvisation teaching, because they offer more options for the creative thinking to develop. The bigger amount of elements for inspiration, the more coherent and effective is the communication through improvisation.

I consider that inter-artistic initiatives in art schools are very effective on reaching children's motivation and developing their imagination and expression. They begin expressing their individuality but, at the same time, they accept and show interest in other individualities within their social group. In a multicultural society, we need to make possible to younger generations to learn in a holistic way, in order to respect each other's differences. Creative drama applied to music improvisation shows great potential in helping children to be aware of their social position in the community and also to stimulate their own critical thinking, preparing for the contemporary world where innovation and creativity are the key to succeed.

6. REFERENCES

Abril, C. (2011). Music, Movement, and Learning. In *R. Colwell & P. R. Webster (Eds.),* MENC Handbook of Research on Music Learning *(Vol. 2: Applications, pp. 93-124). Oxford: Oxford University Press.*

Bandura, A. (1986). Social Foundations of Thought and Action: A Social Cognitive Theory. *Englewood Cliffs, N.J.: Prentice-Hall.*

Bernhart, W. (2006). Narrative Framing in Schumann's Piano Pieces. In *W. Wolf & W. Bernhart (Eds.),* Framing Borders in Literature and Other Media *(pp. 449-476). Amsterdam: Rodopi B. V.*

Boal, A. (1993). Theater of the Oppressed. *New York: Theatre Communications Group.*

Cavadas, L. (2011). Um olhar sobre o meu perfil na orientação de práticas em expressão dramática: um estudo de investigação-ação na Escola Superior de Educação de Viseu. *Mestrado em Arte e Educação, Universidade Aberta, Lisboa.*

Costa, P. (2009). A expressão cênica como elemento facilitador da performance no coro juvenil. Per Musi *63-67.*

Dias, M. (2010). O conto de fadas e a arte de representar. Dois olhares: o do imaginário simbólico e o real da criança. *Mestrado em Literatura Portuguesa - Especialização em Literatura Infanto-Juvenil, Universidade de Trás-os-Montes e Alto Douro, Vila Real.*

Fitzell, G. (2008). Riding the Dragon: Toward a Pedagogy of Free Musical Improvisation. eContact!, 10.4. *Canadian Electroacoustic Community.*

Gersie, A. (1993). On Being both author and actor: Reflections on therapeutic Storymaking. Dramatherapy, 15*(3), 2-11.*

Gonçalves, M., & Trindade, R. (2008). O corpo como lugar da diferença: espaços e tempos de comunicação através da expressão dramática. *Universidade do Porto, in press.*

Hargreaves, A. (1994). Changing Teachers, Changing Times. *New York: Teachers College Press, Columbia University.*

Hugdahl, K. (1995). Psychophysiology: The Mind-Body Perspective. *Cambridge: Harvard University Press.*
Hyry-Beihammer, E. (2011). Narratives in teaching practice: Matti Raekallio as narrator in his piano lessons. Music Education Research, 13*(2), 199-209.*
Kosinski, J. (1992). Selected Essays. *New York: Random House, Inc.*
Leenhardt, P. (1974). A criança e a expressão dramática. *Lisboa: Editorial estampa.*
Levinson, J. (2004). Music as narrative and music as drama. Mind & Language, 19*(4), 428-441.*
Libras, T. (2012). A influência da expressão dramática na qualidade de vida de crianças: um estudo realizado na urbanização Quinta de Lações em Oliveira de Azeméis. *Mestrado em Animação Artística, Instituto Politécnico de Bragança.*
Melo, M. d. C. (2005). A expressão dramática – à procura de percursos. *Lisboa: Livros horizonte.*
Núñez, L., & Navarro, M. (2007). Dramatización y Educación: Aspectos Teóricos. Teoria y educación, 19, *225-252.*
Panofsky, C. P. (2003). 19 - The Relations of Learning and Student Social Class. In Kozulin, A. et al. (eds.) Vygotsky's Educational Theory in Cultural Context. *Cambridge: Cambridge University Press.*
Parncutt, R. (2003). Accents and expression in piano performance. Perspektiven und Methoden einer Systemischen Musikwissenschaft, *163-185.*
Pendzik, S. (2006). On dramatic reality and its therapeutic function in drama therapy. The Arts in psychotherapy, 33*(4), 271-280.*
Sloboda, J. (1991) Music Structure and Emotional Response: Some Empirical Findings. Psychology of Music, *19, 110-120.*
Turpin, G. (1989). Psychophysiology and Clinical-Phychiatry. British Journal of Hospital Medicine, *41(3), 219.*

MUSIC IS COMMUNICATED VISUALLY: MODELING PERFORMER-TO-AUDIENCE COMMUNICATION IN LIVE PIANO PERFORMANCE

Haruka Shoda[1,2], Mayumi Adachi[3]
[1]Department of Culture and Information Science, Doshisha University, Japan, [2]Japan Society for the Promotion of Science, Japan, [3]Department of Psychology, Hokkaido University, Japan
Time: 18:00

ABSTRACT

One of the reasons why an audience experiences powerful emotions in live performance is that the performer communicates his/her performance not only acoustically but also visually. In the present study, we constructed a holistic model of the performer-to-audience communication via acoustical and visual information. Each of 13 pianists performed six pieces (i.e., one fast and one slow piece by Bach, Schumann, and Debussy) in front of 11-23 audience members (N = 211). The duration, the dynamics, and the performer's body movement were measured quantitatively. Both the performers and the audiences rated the affective nuance of each performance by means of 11 adjectives on a 9-point Likert scale. For each performance, the audiences also rated artistry (how artistic the performance sounded), expressiveness (how expressive it sounded), quality (how good it sounded), and their emotionally moving experience (how much the audience was moved emotionally). An exploration using structural equation modeling yielded a statistically valid model with high goodness-of-fit indices (GFI = .96; AGFI = .92; RMSEA = .02). According to the generated model, the compositional features of the piece (i.e., composer, modality, tempo) determined the performer's interpretations of affective nuances of the piece (i.e., valence, arousal), which subsequently determined the durational and the dynamical expressions (e.g., the minor piece induced "negative" and "arousing" nuances in the performer's interpretations, resulting in more variation in dynamics). These acoustical elements also explained the audience's perception of the affective nuance in the same way that the performer expressed. The audience's perception of the artistry and the expressiveness was explained by the degree of his/her perception of dynamics and "arousal," respectively, which subsequently determined the quality and his/her emotionally moving experience for each performance. The performer's body movement did not reflect his/her interpretation but enhanced all of the audience's experiences in listening to the performance. This model captures the complex interactions of the affective nuances of music shared among the composer, the performer, and the audience in the live-performance setting, which makes a further contribution to the existing communication model of music (e.g., Juslin, 2000).

1. INTRODUCTION

"*Music is the language of the emotions*" (Mithen, 2005, p. 24). This *language* can be shared among three agents—the composer, the performer, and the audience—to communicate affective meanings of musical works across time and space. The role of the performer is crucial in this communication: Performers interpret the composer's ideas underlying modality, temporal/dynamical markings, and the structure of the score along with historical background of the piece, and express them through tempo, timing, dynamics, and timbre (e.g., Shaffer & Todd, 1994). The audiences experience the emotions based on these performance parameters (e.g., Juslin, 2000) along with the compositional factors such as modality and melodic progressions (Juslin, 2005). In live performance, musical communication between the performer and the audience is mediated not only acoustically but also visually. Previous studies showed that the visual presentation of music performance enhances the audience's perception of artistry (Shoda & Adachi, 2012b), expressiveness (Broughton & Stevens, 2009; Davidson, 1994; Shoda & Adachi, 2013), quality (Tsay, 2013), and emotions (Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011). Juslin (2000) proposed a model of how the performer's intended emotions are communicated to listeners via expressive performance parameters by using multiple regression. Still unknown is the overall picture of musical communication in the live performance: how the musical language expressed in the score by the composer is communicated to the audience via the performer's interpretation and his/her live performance. The primary goal of the present study was to reveal this mechanism. More specifically, we would extend Juslin's model by using *structural equation modeling* to uncover the complexity of musical communication in the live performance.

Although our attempt is still exploratory, we theoretically hypothesized the overall stream of communications among the three agents. Performers would interpret the affective nuances of the piece ("affective interpretations") by the historical background of composers and the musical features of the piece such as modality and tempo (e.g., Shaffer & Todd, 1994), which subsequently would determine the performer's acoustical expressions (e.g., Juslin, 2000, 2005). The audience would then perceive the performer's affective interpretations as the pianist expressed them (e.g., Juslin, 2000, 2005). Watching the pianist's body movement in a live setting would enhance the audience's evaluations of artistry (Shoda & Adachi, 2012b), expressiveness (e.g., Broughton & Stevens, 2009; Davidson, 1994; Shoda & Adachi, 2012b), affective nuance (e.g., Shoda & Adachi, 2012b, 2013; Vines et al., 2011), and quality (Tsay, 2013).

As for relations between performers' affective interpretations and their body movements, two contrasting hypotheses were drawn. Performers might actively move their bodies to express the affective nuances of the piece as they interpreted, as shown in Dahl and Friberg (2007). Alternatively, performers might intentionally manipulate only "sounds" (Davidson & Correia, 2002; Sandor, 1995) to express such nuances, and their body movements function as artifacts of sound production. In either case, acoustical and visual information would transmit the same information, enhancing the audience's experiences.

2. METHOD

2.1. Participants

Thirteen pianists (4 men, 9 women, 24-40 years old, M = 30.46, SD = 4.41) who held a music degree either in an undergraduate or graduate level participated in this study. Each of them provided one concert featuring the same pieces. They were a concert pianist (n = 1), lecturers at a university or a vocational college (n = 4), piano teachers at private music institutions (n = 7), and a music therapist at a hospital (n = 1). They started to play the piano between ages 4 and 6. Each pianist was paid a 30,000JPY honorarium for participation.

A total of 211 undergraduate and graduate students (98 men, 113 women, 18-59 years old, M = 21.55, SD = 4.50) participated as the audience members, assigned randomly to one of the 13 concerts. Either an extra course credit or a 500JPY honorarium was given as an incentive. The participants' years of private musical training (including composition and improvisation) ranged from 0 to 53 years (M = 10.39, SD = 7.90); 119 of them had experienced piano performance for 1-53 years (M = 10.75, SD = 7.18). We confirmed that the 13 audience samples were drawn from the same population in terms of the age, the year of musical training, and the year of piano training, $F(12, 198) = 1.08$, $p = .38$, $\eta^2 = .06$ (age), $F(12, 198) = 1.46$, $p = .14$, $\eta^2 = .08$ (music training), $F(12, 198) = 1.46$, $p = .14$, $\eta^2 = .08$ (piano training).

2.2. Musical Pieces

We chose six pieces: the b-minor Prelude (Well-Tempered Clavier, Book I, No. 24, BWV869) and the G-major Prelude (Well-Tempered Clavier, Book II, No. 15, BWV884) by J. S. Bach, *Träumerei* (*Kinderszenen*, Op. 15-7) and *Aufschwung* (*Phantasiestüke*, Op. 12-2) by R. Schumann, and *La fille aux cheveux de lin* (*Préludes* Book 2, L. 123-4) and Arabesque No. 1 (Two Arabesques for Piano, L. 66-1) by C. Debussy. We shall refer to these pieces as "B24," "B15," "Dreaming," "Soaring," "Girl," and "Arabesque," respectively. Based on the tempo instruction on the score, faster (B15, Soaring, and Arabesque) and slower (B24, Dreaming, and Girl) pieces were selected for each composer. The modalities of B24 and Soaring are minor, and those of the remaining four pieces are major.

2.3. Procedure

Experiments took place in a small auditorium (with a maximum capacity of 114), equipped with a grand piano (GP-193, Boston) tuned professionally before each concert. Each concert was conducted with a group of 11 to 23 audience members, so that they could have a good view of the pianist.

We asked each pianist to perform the six pieces in the order informed by the experimenter. The pianist performed two pieces (fast, slow) of the same composer as a block. The order of the tempo was consistent across composers. Both the orders of the composer and of the tempo were counterbalanced among the pianists. The performances were recorded onto a multi-track recorder (R24, Zoom) using a microphone (NT4, Rode), and video-taped by two cameras (HF-M32, Canon). A 1cm × 1cm drawing paper marker was attached to the head and waist of each pianist for the measurement of their body movements.

We asked the audience members to attend to both the sound and the pianist for each performance. The audience rated artistry (how artistic the performance sounded), expressiveness (how expressive it sounded), quality (how good it sounded), and their emotionally moving experience (how much the audience was moved emotionally) on 9-point scale, with 1 being "Not at all" and 9 being "Extremely," after each performance. In addition, the audience rated their perceived level of affective nuance of the performance on 9-point scale, with 1 being "Not at all" and 9 being "Extremely," for each of the Japanese equivalents of 11 adjectives (see Figure 1). Each pianist also used the same set of adjectives in rating his/her performance after completing all the performances. At the end of the concert, the pianist and the audience provided demographic information, including the years of musical training and daily experiences in listening to music. The experiment lasted approximately 45 minutes.

2.4. Measurements

From the digital recordings of each performance, the first author measured the duration of each beat (i.e., a dotted quarter note for "Soaring," a quarter note for the remaining five pieces) using Wavosaur (Wavosaur Team). The A-weighted sound pressure level of each performance was measured by using a 1/3 octave band analysis (DSSF 3.5.1, Yoshimasa Denshi). A pianist's body movement was measured by tracing each marker on the head and the waist at the rate of 60.00 fps on a Windows 7 computer using a three-dimensional video analyzer (Frame DIAS IV, DKH). We calculated the pianist's postural angle θ (rad) determined by the locations of head and waist of the pianist (see Shoda & Adachi, 2012a, Figure 2 for the detail). In order for the sound pressure level (dBA) and the postural angle

(rad) to be synchronized with the beat, we calculated "dynamic range" and "movement amplitude," respectively. Each of these values was computed by the difference between the maximal and the minimal values per beat.

For the pianist's acoustical and bodily expressions, we calculated the mean values, the coefficient of variation (CV), and the cross-sectional variation (CSV) of the beat duration, dynamic range, and movement amplitude within each piece. CV is the standard deviation normalized by the mean tempo, indicating overall variability of the target parameter throughout a piece. For CSV, we first calculated a range (i.e., a difference between the maximal and the minimal values) of each parameter within each section of the piece, and then obtained the standard deviation of the ranges as CSV. Thus, CSV indicates the relative variability of the target parameter across sub-sections of a piece.

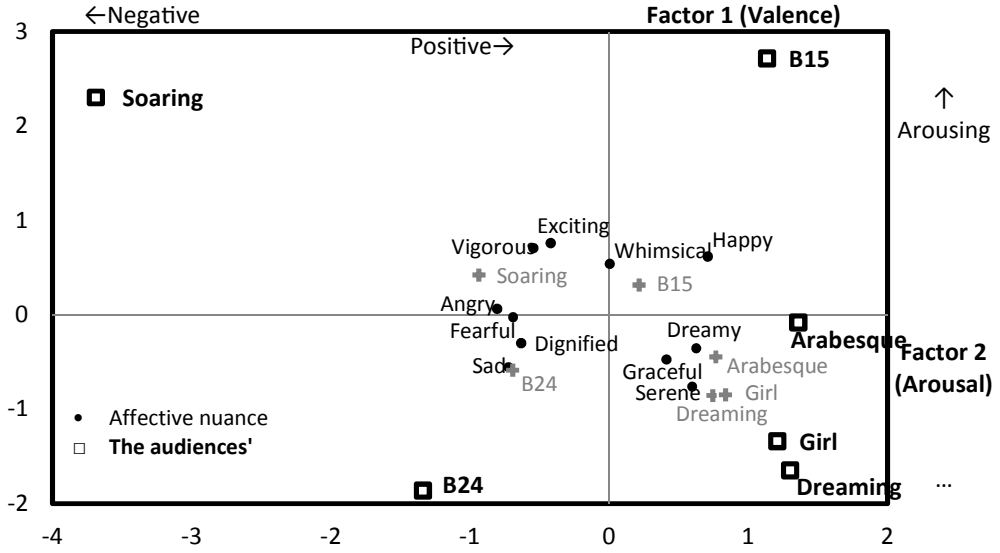


Figure 1: The positioning map of the impressions by the pianists and the audiences, generated by the three-mode positioning analysis (Shoda & Adachi, 2012b). The representative value for the pianist's interpretation was computed by the centroid of his/her ratings, and that for the audience's perception was equivalent with the structured mean computed by this analysis.

2.5. Data Reduction of the Pianist's and the Audience's Ratings

Prior to the main analysis for the model, we first explored data reduction of complex affective nuances of each piece by identifying the structure of the pianist's and the audience's impressions of the six pieces by using a three-dimensional positioning analysis (Shoda & Adachi, 2012b; Toyoda, 2001). Figure 1 shows the structure of the relationships between 11 adjectives (i.e., affective nuances) and the audience's and the pianist's ratings. The positions of the adjectives appear to represent *valence* (positive-negative) as Factor 1 and *arousal* (arounsig-calming) as Factor 2, in line with the dimensional theory of emotions (Russell, 1980). This two-dimensional plane reveals the structure of the pianist's and the audience's impressions for each piece. B15 was located on Quadrant I, indicating that B15 was perceived as a happy and whimsical piece (i.e., positive and arousing) for both the pianists and the audiences. Similarly, Soaring sounded exciting, vigorous, and angry (i.e., negative and arousing); B24 sounded fearful, dignified, and sad (i.e., negative and calming); and the remaining three sounded serene, graceful, and dreamy (i.e., positive and calming). For the following analysis, we used the coordinates (x, y) calculated for each pianist and each audience member for each piece, where x represents arousal interpretation (pianist) or arousal perception (audience) and y represents valence interpretation (pianist) or valence perception (audience).

2.6. Structural Equation Modeling

To understand the "causal" relationships among the composer, the performer, and the audience we applied structural equation modeling (Jöreskog, 1973) to the present data. Structural equation modeling ("SEM") is a statistical technique for testing and estimating causal relations using a combination of statistical data and qualitative causal assumptions (Jöreskog, 1973). This analysis allows us to reveal not only the relationships among the obtained data ("manifest variables") but also those among the covert variables that are not observable but existent theoretically ("latent variable"). For example, we can identify different aspects of "dynamic range" separately through its mean, coefficient of variation, and cross-sectional variation, but we cannot grasp the overall dynamic expression as a whole. The SEM can integrate three separate variables (i.e., mean, CV, CSV) of an acoustical measure (e.g., dynamic range) into one latent variable (such as "dynamic variability" for dynamics), and can illustrate the complex relationships between the manifest and the latent variables. We conducted this analysis by the maximal likelihood method (ML), due to our sufficient sample (N = 211) in the present study.

Generally, a SEM is applied to between-subjects data, so we needed to construct our dataset (i.e., participants (N = 211) × pieces (l = 6)) accordingly. The new datasets were created as follows: First, we assigned each of the 211 participants to one of the 6 pieces (ns = 35-36), and regarded these data as between-subjects, as if each participant was exposed to only that piece. We then confirmed the validity of this assignment by testing 6 different combinations of datasets.

The parameters applied to the SEM were as follows: three composer's variables (i.e., modality, tempo, composer), two affective nuances interpreted by the performer ("interpretation," i.e., valence, arousal), three representative values (i.e., mean, coefficient of variation, cross-sectional variation) for each of the beat duration (i.e., temporal unit), the dynamic range, and the movement amplitude, two affective nuances perceived by the audience ("perception," i.e., valence, arousal), and four variables evaluated by the audience (i.e., artistry, expressiveness, quality, emotionally moving experience).

3. RESULTS

The generated model consisted of four parts (Figure 2): (a) the compositional factors (filled boxes), (b) the pianist's affective interpretations (dark gray boxes), (c) the expressive performance parameters (light gray boxes and ovals), and (d) the audience's experiences (unfilled boxes). The manifest (observable) variables are expressed as boxes and the latent (unobservable) variables as ovals. The model shows an excellent fit for the current data (GFI = .96; AGFI = .92, RMSEA = .02, χ^2 (80, N = 211) = 83.88, p = .36). The paths shown in Figure 2 were all significant (p < .05) except the path from the "movement variability" to the "artistry" (p = .07). The value indicated with each arrow indicates the standardized coefficient (regression weight), which ranges from -1.00 (negative influence) to 1.00 (positive influence). The two latent variables—"dynamic variability" and "movement variability"—were revealed as determinants of the pianist's manipulations of dynamics and body movement, respectively (see Figure 2c), serving as mediators among compositional factors, the pianist's affective interpretations, and the audience' experiences. No latent variable was generated for the pianist's durational manipulations so that the coefficient of variation of the duration for each beat remained in the model as a mediator between the composer and the audience's arousal perception.According to this model, all the compositional factors—modality, tempo, and composer (see Figure 2a)—determined the pianist's interpretations of both valence and arousal of pieces (see Figure 2b). For example, a minor and a faster piece induced negative and arousing interpretations in the pianist, and vice versa. Among the compositional factors, the influence of modality was the greatest for the pianist's affective interpretations (βs = .69 and -.49 for the valence and the arousal, respectively). The modality also influenced the audience's *valence* perception (β = .70) in the same way as the pianist's (see Figure 2d). However, the other compositional factors did not directly induce the audience's experiences but they were mediated by the expressive performance parameters (see Figure 2c). As for the expressive performance parameters (see Figure 2c), the durational variation (expressed as Duration CV, i.e., how much the duration of each beat varied throughout a piece) was determined primarily by composer (β = .72), and secondarily by the pianist's arousal interpretation (β = -.16): The durational variation was greater when the piece was written by a composer in a later period and was interpreted as more calming. The dynamic variability (a latent variable determining mean, CV, and CSV of dynamic range), on the other hand, was determined directly by all the compositional factors as well as by the pianist's affective interpretations of the piece: The dynamic variability was greater when the piece was faster (β = -.23), written in major (β = .26) by a later-period composer (β = .52), and when the piece was interpreted as negative (β = -.30) and more arousing (β = .96). The movement variability (a latent variable determining CV and CSV of movement amplitude) was determined solely by the composer (β = .99).

The audience's affective perceptions (see Figure 2d) were determined in the same way as expressed by the pianist. When the pianists interpreted the affective nuance of a piece as more negative and arousing, the dynamic variability of their sound became higher and induced a more negative (β = -.37) and arousing affect (β = .80) in the audience as well. Similarly, when the piece was interpreted as more arousing by the pianists, the durational variation of the piece decreased, and influenced the audience's perception of the piece as more arousing (β = -.19).

Unlike their valence and arousal perceptions, the audience's impressions of the performance were barely influenced by the acoustical parameters. The only exception was the audience's impression of artistry, influenced by the dynamic variability (β = .20): The higher the dynamic variability was, the more artistic the performance sounded to the audience. Instead of acoustical parameters, every aspect of the audience's experiences was influenced by the movement variability (see Figure 2d). When the pianist's movement variability was higher, the audience perceived the piece as more positive (β = .18) and more calming (β = -.42), evaluated the performance as more expressive (β = .24), more artistic (β = .13), better (β = .16), and more emotionally moving (β = .19).

As for relations between the audience's perception of affective nuances and their impression of the performance, the more calming the performance sounded, the more expressive it sounded to the audience (β = -.26). The degree of expressiveness subsequently influenced the audience's impression of how good the performance sounded (β = .29) and how much they were moved emotionally (β = .42). The last two impressions—the quality (β = .23) and the emotionally moving experience (β = .17)—of each performance were also influenced by how artistic it sounded.

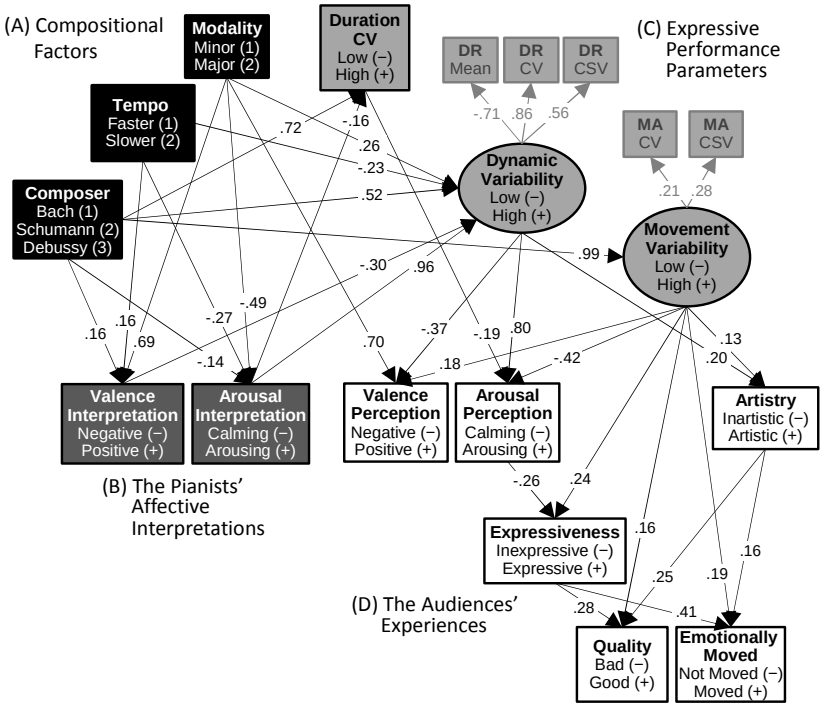


Figure 2: The structural equation model among the composer, the performer, and the audience. The items in the rectangle and the oval indicate the manifest (observable) and the latent (unobservable) variables, respectively. The compositional factors (a) are either the nominal (modality, composer) or the ordinal (tempo) scales: The number in parenthesis indicates the value applied to the analysis. All the other variables were either the interval (ratings) or the ratio (acoustical and physical measures) scales, in which the direction was indicated by the sign (-/+). The abbreviations are as follows: DR (dynamic range), MA (movement amplitude), CV (coefficient of variation), and CSV (cross-sectional variation).

4. DISCUSSION

In the present study, we investigated a possible mechanism of musical communication in a live concert by identifying the causal relationships among the multiple parameters of musical composition, the performer, and the audience, involved in performances of six pieces by Bach, Schumann, and Debussy. According to the present model (see Figure 2), one or more of the compositional features influence the pianist's affective interpretations, the pianist's acoustical and bodily manipulations, and the audience's perception of valence. As predicted, the audience's perception of the affective nuances of the piece was determined by the compositional (i.e., modality) and the acoustical (i.e., durational, dynamical) factors. The acoustical cues, by which the performer expressed valence and arousal of each piece, can be decoded by the audience, in line with the model proposed by Juslin (2000, 2005).

The present model indicates that the pianist's body movement, primarily determined by the composer's historical background, influences the audience's overall experiences including their perception of affective nuances, artistry, expressiveness, and quality, as well as their emotionally moving experience, as predicted from the literature (Broughton & Stevens, 2009; Dahl & Friberg, 2007; Davidson, 1994; Shoda & Adachi, 2012b, 2013; Tsay, 2013; Vines et al., 2011). New here are detailed relationships among the performer's expressive performance parameters and different aspects of the audience's experience. The audience's perception of artistry is not influenced by how affectively or expressively the performance sounds, but by how the performer manipulates dynamics and his/her body movement. The perceived quality of a performance, i.e., an index of the audience's preference (Clarke, 1993), and the emotionally moving experience, i.e., the audience's general emotional responses (Scherer, Zentner, & Schacht, 2001-2002), can be determined by the degrees of artistry and expressiveness perceived by the audience, though the quality of the performance and the audience's emotionally moving experience are not related to each other. According to Scherer et al. (2001-2002), listeners report "being moved" as their affective reactions to classical music more frequently than feeling "basic" (e.g., happy, sad) or "fundamental" (i.e., arousal, valence) emotions. The present model indicates the path from the audience's arousal perception to their emotionally moving experience via their perception of expressiveness. Since this "arousal" perception is generated statistically from the audience's conscious evaluations of affective nuances of a performance (see Figure 1), the arousal perception in this model should be considered as an unconscious process. In other words, even if the audience is unaware of the arousal level of a performance, their perception of its affective nuances would lead to their emotionally moving experience. The present model suggests that even when the audience's "basic" or "fundamental" emotions (Scherer et al., 2001-2002) are not stated explicitly, they can still underlie their general reaction to classical music, "being moved."

The fact that the pianist's affective interpretations do not influence his/her body movement appears to suggest that the primary role of the body movement for the pianist is to produce intended sound (Davidson & Correia, 2002; Sandor, 1995; Shoda & Adachi, 2012a). In addition, the present model suggests that the *variations* (i.e., CV, CSV) of the pianist's body movement are more important for the pianist and the audience than its *size*, as evident in the lacking path from movement variability (i.e., the latent variable for body movement) to the mean movement amplitude. This is in line with Shoda and Adachi (2013), in which the audience appears to determine their evaluations of affective nuances in accordance with the *variation* of the pianist's body movement. The importance of the *variation* in an expressive performance parameter is also evident in the overall variability of the duration of beat (i.e., Duration CV in Figure 2c), the only variable of the durational manipulation identified as functional in this model.

In conclusion, we identified roles of the music composition, the performer, and the audience in musical communication during live piano performance by using structural equation modeling. Even though the present model has demonstrated the holistic picture of musical communication in a live context, it has not uncovered the detail of the pianist's intended sound underlying his/her body movement, as evident in the lack of relationships among durational, dynamic, and movement variations. To reveal them, time-series analyses should be conducted in the future. Moreover, the external validity of this model needs to be tested with other pieces of the same composers, pieces of other composers, and with other pianists.

5. ACKNOWLEDGMENT

The experiments were conducted as a part of the first author's doctorate study at Hokkaido University, Japan, supported by Grant-in-Aid for JSPS Fellows (10J00985). We thank George Waddell for his proofreading of this manuscript. We are grateful for all the pianists and the audience members for their participations in the present study. We thank Kenji Watanabe, Kenichiro Takahashi, Shoko Fukai, and Ayumi Inoue for their supports in recruiting the experimental participants. We also thank Nami Koyama, Noriko Ito, Yosuke Tani, Kazuma Takiuchi, Chihiro Suzuki, Ayumi Sasaki, Ding Xingxing, Huo Xinyang, and Chen Lingjing for their assistances in the experiments.

6. REFERENCES

Broughton, M., & Stevens, C. (2009). Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. Psychology of Music, 37*(2), 137-153.*

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. Music Perception, 10*(3), 317-341.*

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musician's body movements. Music Perception, 24*(5), 433-454.*

Davidson, J. W. (1994). Which areas of a pianist's body convey information about expressive intention to an audience? Journal of Human Movement Studies, 26, *279-301.*

Davidson, J. W., & Corriea, J. S. (2002). Body movement. In R Parncut, & G. E. McPherson (Eds.), The science and psychology of music performance: Creative strategies for teaching and learning *(pp. 237-250). New York: Oxford University Press.*

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger, & O. D. Duncan (Eds.), Structural equation models in the social sciences (pp. 85-112). New York: Seminar Press.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. Journal of Experimental Psychology: Human Perception and Performance, 26*(6), 1797-1813.*

Juslin, P. N. (2005). From mimesis to catharsis: Expression, perception, and induction of emotion in music. In D. Miell, R Macdonald, & D. J. Hargreaves (Eds.), Musical communication *(pp. 83-115). New York: Oxford University Press.*

Mithen, S. J. (2005). The singing Neanderthals: The origin of music, language, mind and body. *London: Weidenfeld & Nicolson.*

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39*(6), 1161-1178.*

Sandor, G. (1995). On piano playing: Motion, sound, and expression. *New York: Schirmer.*

Scherer, K. R., Zentner, M. R., & Schacht, A. (2002). Emotional states generated by music: An exploratory study of music experts. Musicae Scientiae, 6*(1), 149-171.*

Shaffer, L. H., & Todd, N. P. M. (1994). The interpretive component in musical performance. In R. Aiello, & J. A. Sloboda (Eds.), Musical perceptions *(pp. 258-272). New York: Oxford University Press.*

Shoda, H., & Adachi, M. (2012a). The role of a pianist's affective and structural interpretations in his expressive body movement: A single case study. Music Perception, 29*(3), 237-254.*

Shoda, H., & Adachi, M. (2012b). Effects of the listening context on the audience's perceptions of artistry, expressiveness, and affective qualities in the piano performance. In E. Cambouropoulos, C. Tsougras, P. Marvromatis, & K. Pastiadis (Eds.), Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music *(pp. 925-929). Thessaloniki, Greece: ICMPC12.*

Shoda, H., & Adachi, M. (2013). Expressivity, affective nuance, and presentation modality in a performer-to-audience communication. *Manuscript submitted for publication.*

Toyoda, H. (2001). Tansaku teki positioning bunseki: Semantic differential data no tame no 3 so tahenryo kaisekihou [An exploratory positioning analysis: Three-mode multivariate analysis for semantic differential data]. The Japanese Journal of Psychology, 72*(3), 213-218.*

Tsay, C. J. (2013). Sight over sound in the judgment of music performance. Proceedings of the National Academy of Sciences of the United States of America, 110*(36), 14580-14585.*

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. Cognition, 118*(2), 157-170.*

[2C] Musicking

MONTREAL 16:30-18:30 Monday 4 Aug 2014

DELAYED AUDITORY FEEDBACK DISRUPTS IMPROVISED PIANO PLAYING MORE THAN MEMORIZED PLAYING: TOWARDS A COGNITIVE CHARACTERIZATION OF IMPROVISATION

Andrew Goldman[1]
[1]*Centre for Music and Science, University of Cambridge, United Kingdom*

Time: 16:30

What cognitive features characterize improvisatory music performance and distinguish it from other modes of performance such as playing from memory? Delayed auditory feedback (DAF) has been used to make inferences about perceptual-motor processes underlying piano playing based on its disruptive effects. Because improvisation may engage motor planning systems differently than rehearsed performance, might DAF help distinguish and characterize improvisatory cognitive processes?

Generally, this research aims to help develop a theoretical approach to characterizing improvisation based on comparing its perceptual-motor processes with other modes of performance rather than focussing on questions associated with how performers generate particular musical structures. Specifically, it does so through looking at the differing effect of DAF on improvised compared with memorized playing. 10 jazz pianists played walking bass lines on a MIDI keyboard over a single chorus of Rhythm Changes in a 2x2x2 fully factorial design, four trials per condition. The bass lines were either memorized or improvised, with or without DAF, and in B or Bb. Participants were instructed to play only crochets, one note per beat. The memorized bass line was given to participants beforehand to learn. Before playing each bass line, participants played an arpeggio with a metronome for 8 bars to set an identical tempo for each trial. After this, the metronome dropped out. For trials with DAF, delay was implemented only after this synchronization period. The average inter-onset interval (IOI) and the IOIs coefficient of variance were calculated for the post-metronome portion of each trial. A three-way repeated measures ANOVA was conducted on each metric. For both metrics, there was a significant main effect of performance mode; improvised trials had larger average IOIs and a higher coefficient of variance than memorized ones. As would be expected from previous research, there was a significant main effect of delay; DAF trials had higher IOIs and coefficients of variance. Importantly, there was a highly significant interaction between performance mode and delay. DAF raised average IOIs and the coefficients of variance more for improvised trials than memorized ones. Improvisation was more disrupted by DAF than memorized playing. In this musical context, differences in improvisation's underlying perceptual-motor processes have been demonstrated, suggesting musicians rely more on auditory feedback while improvising. This finding is situated within discourses of improvisers learning to mentally 'hear' what they play and contributes to a cognitive characterization of improvisatory motor processes.

PERCEPTIONS OF LEADERSHIP IN DUO KEYBOARD IMPROVISATIONS

Freya Bailes[1]*, Roger Dean*[2]
[1]*University of Hull, United Kingdom,* [2]*MARCS Institute, University of Western Sydney, Australia*

Time: 17:00

Enacting different leadership roles in ensemble improvisation necessitates the perception of and response to another player, leaving an audible record of these cognitive processes. Musical transitions are of particular interest in solo and group improvisation. At such moments, cognitive effort is required to break an old pattern and develop a new pattern, with an observable increase in attention at points of change. We aimed to explore the physiological and perceptual correlates of musical change in duo improvisations, to uncover processes of leadership involved in the introduction and evolution of audible musical ideas. Six pairs of professional keyboard improvisers were invited to perform a series of improvisations, sitting back-to-back at MIDI keyboards. MIDI, audio, video, and skin conductance were recorded during the improvisations. The first four improvisations were to be free or based around a referent. The next two improvisations attributed a leadership role to each improviser of the pair. The improvisers subsequently listened back to a subset of their improvisations, with the audio balanced to hear one player in each ear. While listening, they continuously moved a cursor along a left-to-right scale to reflect the musician they felt was most influencing the musical progression. Computational analyses of the MIDI data were conducted to identify segments and transitional passages in the improvisations. Time series analysis tested the relationship between musical structure, performers' perceptions of changes in musical leadership, and their physiological arousal while playing. Players were largely in agreement as to the moments that each had led during the improvisations. Substantial differences were found between their signature skin conductance patterns. Nevertheless, temporal alignment was found between increases of the performers' physiological arousal and their exhibited and perceived musical leadership. Aural cues were sufficient for substantial agreement between improvisers with respect to the time course of perceived leadership, but complex statistical relationships in the data reflect the complexity of the many intermediary variables that can be found in ensemble improvisation.

DONATONI: A CRITICAL RE-APPRAISAL OF QUARTETTO III

Massimo Avantaggiato[1]
[1]*Conservatorio G. Verdi di Milano, Italy*

Time: 17:30

This paper provides an analysis of Quartetto III, composed by the Italian musician Franco Donatoni. It is Donatoni's third Quartet (Donatoni wrote six quartets; the last one was La Souris Sans Sourire - 1988), Quartetto III, unlike the other

quartets, was produced, under the guide of Marino Zuccheri, at the "Studio di Fonologia" in Milan, by using only electronic instruments. We've studied the musical, historical and technological context in which this work was conceived by using different historical sources. This paper highlights some specific ideas of Quartetto III, the composer's originality and importance not only in his work, but also in the the production of the Studio of Fonologia and in the history of electroacoustic music. In the 1950s, while Boulez, Stockhausen and Berio were investigating electronic music and total serialism, Donatoni was writing in a comparatively safe neo-classical style, with a strong initial influence by Bartok. In 1953 Donatoni met Bruno Maderna, who introduced him to Webern's music and the European avant-garde. Donatoni and Maderna will be two of the twelve composers that will work at the Studio di Fonologia in Milan during the 60s. Quartetto III has been analysed under different points of view:We have pointed out the relationship between ministructure and macroform, underlining the progressive aggregation process , from "elements" to "Groups" and "Columns". This objective has been achieved by means of: a) a Genetic Analysis by using PWGL; b) a listening Analysis, by following different musicological approaches: Pierre Schaeffer; Chion and Delalande;

- Spectro-morphology;

- Roy Functional Analysis;

- Sloboda, McAdams: Perceptive and cognitive Studies.

This approach can give us some information about the macrostructure: Quartetto III, which lasts about 5 minutes, is structured in panels (sections with different metronomes, but with an internal coherence of articulation and musical development. Much attention has been paid to polyphonic structure and quadraphonic space. Quartetto III seems to pave the way for the later electroacoustic works because of the use of spatial figures and "structured" electronic gestures.

RHYME, REASON, AND RHYTHM: ELISION, ENJAMBMENT AND ENTROPY IN THE PHRASING AND RHYME SCHEMES OF RAP

Nathaniel Condit-Schultz[1]
[1]*The Ohio State University, USA*
Time: 18:00

ABSTRACT

Rap consists of highly stylized through-composed deliveries of poetic lyrics in a sing-song or chant like style. Several theoretical papers have discussed the musicality and creativity of rap 'flow' (Adams 2008,2009 ; Ohriner 2013). The interaction between the meter, rhyme schemes, musical as well as semantic/slash grammatical phrasing, and surface rhythms, are what generates musical interest in rap. The current paper presents a systematic empirical analysis of rap flow in a large corpus of popular rap works. A corpus of over 200 transcriptions of rap songs by 30 different rappers is encoded in a modified Humdrum encoding. Transcriptions include representations of rhythm, syllable stress, rhyme, and phrase boundaries (both musical and semantic). Important rhythmic events in flow are quantified, notably the metric placement of rhymed syllables and phrase boundaries. Information theory measures are used to demonstrate the historical trend towards greater complexity in rap, and to make empirical comparisons of complexity in the styles of different rappers. The zeroth-order probabilities associated with the metric positions of rhymed syllables are counted and larger scale patterns, particularly first-order relationships between consecutive rhymes, are investigated. Final results awaits the completion of the corpus.

1. INTRODUCTION

Over the last three decades hip-hop music has established itself as a major genre in western popular music, much as rock did 30 years earlier. Hip-hop's short and well documented history offers a unique opportunity to observe the development of a musical genre from its genesis to the present. The most novel and unique feature of hip-hop is its definitive style of vocal performance: "rap." Rap consists of a highly stylized, through-composed, delivery of poetic lyrics in a sing-song or chant-like style. By deemphasizing discrete pitches compared to other types of song, rap brings rhythmic and poetic structures to fore as the focus of artistic expression. The rhythm, phrasing, and poetics of rap delivery are often referred to as "flow." Like other musical structures, rap flow balances the expected with the surprising, creating and releasing tension. This paper presents a theoretical outline of the artistry of rap flow, and describes a corpus based project intended to inform and test this theory, as well as characterize and quantify the qualities of various rap performances.

For listeners unfamiliar with rap, it is often difficult to recognize the musicality of rap "flow." The nature of rap lyrics—which are obtusely self-referential, ensconced in dense slang and obscure sub-cultural references, and often extremely offensive—has discouraged many listeners from making any effort to appreciate rap's musicality. Thus, rap's status as art, or even as music, is questioned and denigrated more so most other genres. By further elucidating the musical sophistication of rap, we hope to spread an appreciation and recognition of rap as a musical art.

Since the poetic nature of rap is more apparent than its musicality, most scholarly (and lay) discussion of rap has approached it as poetry. However, neglecting the musical aspects of rap delivery leads to an impoverished understanding of the art. In particular, poetic analysis of rap typically fails when trying to use concepts of poetic meter to describe rap. As an example, consider the delivery of the song "The Way I Am" by the rapper Eminem. Several Wikipedia pages describe the verses of this song as being in 'anapestic tetrameter,' a term from poetic meter referring to syllables organized into poetic feet in a weak-weak-strong pattern. However, this analysis neglects the most crucial feature of this delivery, that this weak-weak-strong pattern is syncopated in relation to the musical meter. Eminem delivers these feet on the 2nd, 3rd and 4th sixteenth-notes of each quarter-note beat. Theories of poetic meter are not adequate to capture this musical fact. This example should illustrate that rap delivery must be understood in terms of musical meter, not poetic meter. Poetic

meter does not refer to an independent hierarchy of isochronous pulses and sub-pulses, as musical meter does. Other mainstays of poetic traditions, such as counting the number of syllables, feet, or mora in a given line are also irrelevant to rap as well, since the independent musical accompaniment provides a metric hierarchy against which rappers have greater freedom to vary their flow. Finally, it must be emphasized that unlike many poetic traditions rap is not a written art form; the concept of 'lines' in poetry, explicitly delineated by carriage returns in a written document, does not have a literal analog in rap.

As with most forms of vocal music (and poetry), rap combines musical structures with semantic meaning. The meaning of the lyrics, including the use of metaphor, story-telling, and political/social messages, is clearly an important part of many song genres, rap included. What's more, the musical and lyrical subject matter may interact, creating an emergent artistic product which is greater than the sums of its parts ('text setting'). The current study is limited almost entirely to analysis of the musical aspects of rap flow, without reference to the semantic content of the lyrics. This does not reflect a value judgment about the relative importance of the two elements, nor is it a dismissal of the interaction between the two. Rather, the approach simply reflects a standard empirical 'divide and conquer' reductionist method, seeking a complete understanding of the parts before attempting to explain the emergent whole. In defense of the current approach, we can point to numerous examples of songs, including rap songs, which are appreciated and enjoyed by listeners who do not understand the lyrics, or even the language of the lyrics. These cases support the notion that the purely musical structures of rap can stand on their own as an artistic product. We also do not discuss the musical accompaniments to the raps.

2. LITERATURE REVIEW

A number of commercial books intended for popular audiences have been written about rap and rap flow. The majority of writing on rap, both scholarly and lay, has focused on the social issues surrounding hip-hop culture and rap lyrics, from the perspective of communication studies and musicology. Noteworthy books which do consider the artistry of rap delivery itself include Adam Bradley's "Book of Rhymes: The Poetics of Hip Hop" and Paul Edwards' "How to Rap: the Art and Science of the Hip-Hop MC." In a more scholarly vein, the most significant musical analysis of rap can be found in two papers by Kyle Adams (2008; 2009). The current project is largely inspired by Adams' work. In terms of empirical analysis, the most important precursor to this paper is the dissertation work of Hussein Hirjee (2010). Hirjee, conducted a corpus analysis of texts from several thousand rap songs by twenty five popular artists. Hirjee noted a marked increase in the complexity of rhyme patterns found in rap from its roots in the early 1980s through the year 2000. However, Hirjee's analysis, doesn't included any rhythmic information, approaching rap songs essentially as written texts.

3. THEORY OF RAP 'FLOW'

A typical rap verse consists of a nearly continuous delivery of words over 16-32 measures of music (nearly always 4/4 time). Verses consist of a series of short phrases of varying length, typically one phrase per measure of music, with each phrase roughly corresponding to a single sentence or clause. How the phrases of a rap are set to the meter, what types of rhythms, and the number of syllables per line are all highly variable. Raps are structured by 'sonic parallelisms' both across and between phrases. 'Sonic parallelism' refers to the partial or complete repetition of aspects of the speech signal. The most obvious parallelisms in rap are rhymes, but other elements of speech can create notable parallels as well. To refer broadly to all sorts of 'sonic parallelism' in rap delivery I propose the term 'chyme.' Chyme can include phonemic parallels (i.e. rhyme), prosodic parallels (melodic/intonation figures), rhythmic parallels, and parallels involving timbral or articulatory effects. Very often it is the case that a particular chyme will include several or all of these features.

Chymes delineate phrase and higher level groups, and also serve as bridges and connections between and across phrases. In many cases, rappers use chymes in a regular, predictable, manner creating clear 'chyme schemes' (what Ohriner refers to as "grooves"; 2013). In these cases, the expectations of chyme serve as important creators and releasers of tension. Prototypically, an 'end chyme' appears at the end of each phrase. In most rap, the broad expectation of the arrival of a chyme to parallel the ending of the previous line gives the music its sense of forward momentum. Other chyme schemes can be more complicated.

For instance, consider the opening eight measures of Eminem's "Drug Ballad" (Figure 1). The phrases in this passage are each one measure in length and are built around a three-part chyme scheme: A three-syllable end chyme landing on beat three (rhyme and rhythm), a syncopated anucrusis just before each down beat (rhythmic chyme), and a two-syllable chyme which answers the anucrusis chyme.

The theory proposed here is that the musical interest of rap is created by the interaction of prosodic phrasing, semantic phrasing, sonic parallelism (chymes), and the manner in which these structures are delivered in relation to the musical meter. Rappers synchronize these structures in predictable patterns (chyme scheme). However, rappers rarely repeat a chyme scheme for an extended time without variation, and frequently morph from one scheme to another as their verses progress. In some cases, the alignment between the various structures is offset or varied, as when a semantic phrase (such as a sentence) ends with the first syllable of a new musical phrase (similar to enjambment in poetry). The aim of this project is to analyze rap deliveries in respect to the model outlined above, in particular to characterize and quantify the complexity, and predictability of rap passages, using information theory measures such as entropy. By adopting a corpus approach, the goal is to characterize passages not in isolation, but in relation to the average, 'norms,' of the genre as a whole.

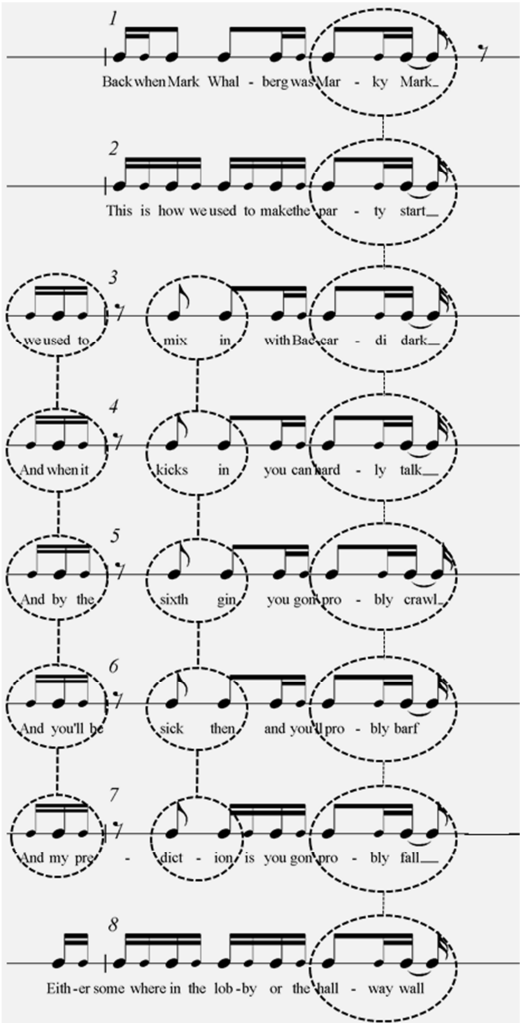


Figure 1: The first eight measures from Eminem's "Drug Ballad." Each line represents one measure of music. Dashed circles connect chymed syllables.

4. CORPUS

As is often the case in arts research a representative sample of rap is not necessarily desirable. What interests us, especially when our resources limit us to a small sample, are the exemplars of the genre: artists and works known to be of the highest quality. To avoid bias towards artists familiar to the author, an independent source was sought to offer criteria for inclusion in the sample. Sampling was restricted to rappers and songs that have appeared on the Billboard Top 100. The advantage of this approach is that the Billboard chart does not reflect the biases or opinions of a single individual, but rather the collective choices of millions of consumers. We presume that commercially successful artists differ from unsuccessful artists at least partly because they are 'better' artists. Of course, successful artists may also differ in other ways which are not as interesting (better marketing perhaps) as well, and singles may differ systematically from the repertoire as a whole. Still, we can at least be confident that Billboard charts represent the most widely listened to music. Despite reservations, Billboard presents itself as the best source for an unbiased, systematic sample of rappers, and rap songs.

1,305 rap songs charted on the Billboard Top 100 between 1980 and January 2014. Even restricted to the Billboard data, there are many possible ways to operationalize charting success. For instance, an artist might be rated by the number of songs to hit number 1, or to enter the top 40, or by the number of weeks their songs remained on the chart. Our approach was decided to rank artists simply by the number of their singles to have appeared on the chart. The thirty-three artists sampled, in order from the most charting singles to least, are: Jay-Z, Eminem, Lil Wayne, Kanye West, T.I., Nelly, Ludacris, 50 Cent, LL Cool J, Pitbull, Snoop Dogg, 2pac, Bow Wow, The Black Eyed Peas, Busta Rhymes, Will Smith, OutKast, T-Pain, Fabolous, Fugees, The Notorious B.I.G., Nas, Rick Ross, Fat Joe, Ja Rule, Puff Daddy, Young Jeezy, DMX, Salt 'N Pepa, M.C. Hammer, Missy Elliott, Naughty By Nature, and Twista. The initial sample consists of the five highest charting single from each artists' first five years on the chart.

5. TRANSCRIPTIONS

Transcriptions were created in a modified Humdrum text format (Huron 1999). A text file is created with each tab delineated column ("spines" in Humdrum lingo) representing a different type of transcribed information, and time represented as going downward in the file. Lyrics were translated to an IPA (International Phonetic Alphabet) representation using a program developed by the author. All translations were carefully examined and corrected by the author, to represent as closely as possible the actual pronunciation in the recording.

In transcriptions, the syllable was taken as the rhythmic unit. Each syllable was placed on its own record in the text file. Each record represents a metric subdivision, usually a sixteenth-notes, but triplet sixteenth-notes and thirty-seconds are included where needed. The metric position of each record was indicated in the first spine (column) of the text document. Rap deliveries feature many subtle rhythmic features (such as rapping "behind the beat") which are not captured by this encoding scheme. Consideration of the effect of rhythmic nuances is left to future research.

In addition to the syllable pronunciation and rhythm, transcriptions include codings of syllable stress, prosodic phrase boundaries, semantic phrase boundaries, and the location of rhymes.

6. ANALYSES

The first analytical goal is to characterize the 'norms' of rap. In one body of analyses, metric time is taken as the reference unit: The occurrence of syllables (stressed and not), chymes, as well as phrase onsets and endings are counted for each position in the metric hierarchy. The number of syllables, rhymes, and phrase boundaries per measure are also counted. In a second set of analyses, the phrase, as delineated by prosodic boundaries included in transcriptions, is taken as the primary unit of analysis: Length of phrase (in musical time); metric positioning of phrase (where in the measure the phrase starts and ends); number of syllables in phrase (accented and total), are counted. Use of chymes in phrases is also measured: number of chymes in phrase; locations of chymes within phrase (is the chyming within the phrase, or connected to a previous phrase?) are classified. Finally, lower level rhythmic groups (3-6 syllables in length) are also counted and classified using similar measures.

The goal of these analyses is to create a taxonomy of rhythmic, metric, and phrase patterns employed in rap. The distribution, and relative commonality or rarity, of different figures in the taxonomy can serve as a reference point of zeroth order probabilities in rap in general. Thus, we seek to determine to what degree, and in what features, Eminem's phrases shown in Figure 1 are typical or atypical of rap as a whole.

The next step in the analysis consider repetition and parallelism across phrases. By employing an auto-correlational 'rotation' metric each delivery can be compared against itself rotated at various time scales in order to quantify the amount of, and wavelength, of repetition and parallelism in a given passage. These analyses will reveal the location and complexity of repetitive schemes in the deliveries, allowing characterization of higher-level patterns in a verse and addressing questions such as: Does the rapper use consistent schemes at all? How often does a rapper change chyme schemes? Does he abruptly change schemes or change gradually? Do chyme schemes change at hyper-metrical positions (four or eight measure)? These questions can then inform quantification of higher order probabilities.

How likely (or unlikely) is the fourth phrase of Figure 1, given the three before it?

By characterizing the complexity of different aspects of rappers flow we seek to demonstrate the historical trend towards greater complexity in rap, and to make empirical comparisons of complexity in the styles of different rappers, both from each other and from the norms of the style. A simplified pilot study of a small corpus of thirty songs has revealed the general usefulness of the current approach, including a statistically significant ($p < 0.001$) trend towards greater entropy in the rhyme patterns used by rappers from the 1980s through the 2000s. We will also seek to create analyses of the predictability, and unpredictability, of a rap deliveries as they unfold over time, measuring the ebb and flow of tension and relaxation in the music. Final results awaits the completion of the corpus.

7. REFERENCES

Adams, K. (2008). Aspects of the music/text relationship in rap. Music Theory Online, 14(2). *http://www.mtosmt.org/issues/mto.08.14.2/mto.08.14.2.adams.html*

Adams, K. (2009). On the metrical techniques of flow in rap music. Music Theory Online, 15(5). *http://www.mtosmt.org/issues/mto.09.15.5/mto.09.15.5.adams.html*

Bradley, A. (2009). Book of Rhymes: The Poetics of Hip Hop. *New York, NY: Basic Books.*

Edwards, P. (2009). How to rap: the art and science of the Hip-Hop MC. *Chicago, IL: Chicago Review Press.*

Hirjee, H. (2010). Rhyme, rhythm, and rhubarb: using probabilistic methods to analyze Hip Hop, poetry, and misheard lyrics *(Doctoral dissertation). University of Waterloo, Ontario Canada.*

Huron, D. (1999). Music research using Humdrum: A user's guide. *Center for Computer Assisted Research in the Humanities. Stanford, CA.*

Kats, J. (2008). Towards a Generative Theory of Hip-Hop. Paper. *Paper presented at the Music, Language, and the Mind conference. Tufts University.*

Ohriner, M. (2013). Groove, Variety, and Disjuncture in the Rap of Eminem, Andre 3000, and Big Boi. *Paper presented at the annual meeting of the Society for Music Theory. Charlotte NC.*

[2D] Music and Movement 2

LIÈGE 16:30-18:30 Monday 4 Aug 2014

RHYTHMICAL STRUCTURES IN MUSIC AND BODY MOVEMENT IN SAMBA PERFORMANCE

Mari Romarheim Haugen[1], Rolf Inge Godøy[1]
[1]Department of Musicology, University of Oslo, Norway
Time: 16:30

ABSTRACT

Samba groove is often characterized by its complex rhythmical patterns. Recent studies, based on audio recordings of samba music, report that the 3rd and the 4th 16th-notes are played slightly ahead of their corresponding quantized position, and that this seems to be a prominent feature of samba groove. Considering that samba derives from a culture where music and body motion are intrinsically related, may suggest that we should include both sound and motion data in studies of its rhythmical structures. In this paper we investigate whether the microtiming features, previously shown in samba music, may also be represented in the body motion of performers playing and dancing samba. We report from a motion capture experiment where two skilled samba performers, a percussionist and a dancer, were recorded using an advanced optical infrared motion capture system. Our audio analysis confirms the existence of systematic microtiming patterns on the 16th-note level in samba music. In addition, motion analysis of the percussionist's heel tapping and the dancer's steps revealed motion patterns in synchrony with the systematic microtiming features found in samba music. These observations support the view that the systematic micro timing of 16th-notes in samba playing is not a deviation from an underlying perceived pulse with isochronous subdivisions, but rather constitutes an essential feature of samba.

1. INTRODUCTION

The groove of samba is often characterized by its complex rhythmical patterns, and it is suggested that systematic microtiming is an essential feature of samba groove. In a study of rhythmical structures in Brazilian percussion based on analyses of field recordings (audio) and interviews, Gerischer (2006) documents a prominent medium – short – medium - long duration pattern on 16th-note level, and the existence of the both extended (in duration) and accentuated 4th 16th-note in a beat in particular, related to the production of samba groove. Gerischer refers to the concept of participatory discrepancies (PDs) (Keil 1987) and the hypothesis of systematic variations of durations (SYVAR) in music performance (Bengtsson 1987, Gabrielsson 1982), and emphasizes that the microtiming features in samba groove constitute essential aspects of style. Similar findings are presented in Gouyon's (2007) study of microtiming features in samba. Based on computer analysis of audio excerpts with traditional samba music, Gouyon found that the 3rd and the 4th 16th-notes in samba groove are played slightly ahead of their corresponding quantized position. However, thinking of systematic microtiming in terms of deviations, indicates a culturally biased norm with isochronous time marks as starting point (Bengtsson 1987, Kvifte 2004). In an attempt to sidestep this cultural bias, Kvifte (2004) suggest a *pattern concept*, using "participatory pattern" instead of PDs and "durational pattern" instead of SYVAR. In addition, based on the view that rhythm is intrinsically related to motion, both physical and virtual, (see e.g. Blom & Kvifte 1986, Shove & Repp 1995, Iyer 2002), motions of perceivers and performers may offer a more perceptual relevant reference structure for rhythm studies than an abstract fixed clock pulse (Blom 1981, Bengtsson 1987, Kvifte 2007). Baily (1985) states that human motion is the process through which musical patterns are produced, and emphasizes the need to investigate the relationship between motion and music in the process of music making. In his studies of Norwegian folk music and dance, Blom (1981) takes as his point of departure that musical rhythm is intimately related to our bodily experiences, and that our concepts of rhythm are mirrored by the way in which we move our body in synchrony with music.

Samba music is based on oral traditions and elements of African and Afro-Brazilian rituals, on cultures where music and motion are intrinsically related (Mariani 1998, Carvalho 1999, Gerischer 2006, Naveda 2011). In some African languages, there is no word for music that does not also include the act of dancing (Grau 1983, Baily 1985, Kubik 1990). In a study of Brazilian drum patterns, Kubik (1990) emphasizes that percussionists' inner "elementary pulsation", either objectified by actual strokes or by being silent, serves as the most important temporal orientation for the performers. When transcribing the drum rhythms, Kubik found that the easiest way to identify the performer's "common beat" and also the position of the first beat (inner "reference beat"), was to look at the body motion of performers and dancers.

Considering that samba derives from a culture where music and dance are intrinsically related, suggest that we should include dance motions when studying rhythmical structures in samba. In addition to dance motions, body motion of percussionists may also contribute to a better understanding of rhythmical structures in samba. In the present study we aim to investigate whether the microtiming features, previously shown in the sound of samba, may also be reflected in the body motion of performers playing and dancing samba.

2. METHOD

Two skilled samba performers, a percussionist and a dancer, participated in the study. The percussionist played a samba groove on a pandeiro, a Brazilian hand frame drum with jingles (platinelas). A pandeiro is typically between 10″ and 12″ in diameter with sound in middle to high frequency registers (Naveda 2011). The drum is held in one hand and played on by the other. In our experiment, a small microphone connected to an amplifier was attached to the frame of the pandeiro to augment the low frequent bass sounds from the drum. The dancer performed a dance in samba-no-pé style, an individual dance that can be performed both with and without improvisation (Naveda 2011). The dance was also performed both with and without high-heeled shoes.

The participant's motions were recorded using an advanced optical motion capture system from Qualisys. The system tracked the motions of reflective markers at a frame rate of 200 Hz. All sessions were videotaped using a digital video camera (SONY HandyCam). In addition, sound was recorded using Logic Pro software running on a MacBook computer. A total of 31 reflective markers were attached to each participant's body (Figure 1). The placement of the markers on the percussionist and the dancer were identical.

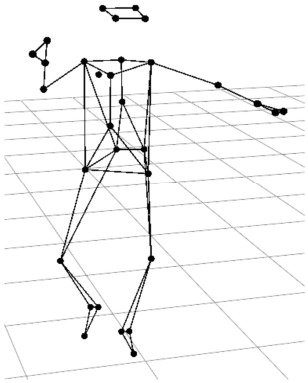


Figure 1: The placement of the markers attached to the participants' bodies.

Ten recordings, with three different setups, were carried out (Table 1). First, we did two recordings of the percussionist alone. Second, we did three solo recordings of the dancer dancing to the music recorded in the second percussion recording. Third, we did five recordings of the percussionist and the dancer together.

Table 1: An overview of the ten recordings made in this experiment. In session I we did recordings of the percussionist alone, in session II we did solo recordings of the dancer dancing to the recorded music from session I, and in session III the percussionist and the dancer where recorded together.

Recordings		Comments
I. Percussionist (solo)	**a.**	Motion capture and sound recording
	b.	Motion capture and sound recording Sound recording is being used in II.
II. Dancer (solo)	**a.**	Samba no pé
	b.	Samba no pé
	c.	Samba no pé with variations
III. Percussionist and dancer	**a.**	Samba no pé with variations
	b.	Samba no pé
	c.	Samba no pé with shoes
	d.	Samba no pé with shoes
	e.	Dance improvisation

3. ANALYSES AND RESULTS

The sound data from the recordings were analyzed using the MIR Toolbox for MatLab (Lartillot & Toiviainen 2007), a toolbox that is developed for extracting musical features like timbre, tonality and rhythm, from audio files. The motion data, obtained from the motion capture recordings, were analyzed using the MoCap Toolbox for Matlab (Burger & Toiviainen 2013). The MoCap Toolbox contains a set of functions for analyzing motion data and is developed for studies of music-related motion. All statistical analyses were performed using SPSS version 21.

3.1. Sound

In the present study, the samba groove was played on a pandeiro. Different stroke techniques on the drum produce different sounds that can be classified as followed: low-pitched bass sound (thumb), high-pitched bass sound (thumb), slap sound (fingers) and jingle sound (fingertips, thumb and palm) (Roy et al. 2007). A typical pandeiro rhythm emerges by alternating between the strokes of the thumb, fingertips and palm. In addition, a quick turn made by the hand holding the drum, makes an accentuated jingle sound.

The samba groove was notated in 2/4 meter. The second beat in a bar is accentuated by a low frequency bass sound, something that is in accordance with previous descriptions (Gerischer 2006, Naveda 2011). In addition, the groove has a four bar periodicity, indicated by a variation played at the end of every fourth bar (sometimes in the end of the second bar). Before the groove starts, a short start break, normally consisting of two 8th-notes, is played. A transcription of the samba groove pattern (simplified) played on pandeiro is shown in Figure 2. The accented jingle sound, caused by the quick turn of the drum, is marked as accents above the notated jingles.

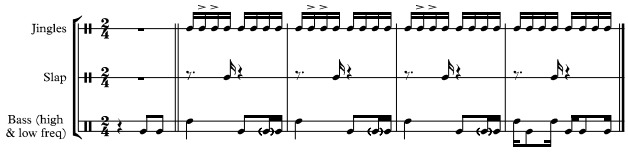


Figure 2: Transcription of the start break and the samba groove played on pandeiro.

Here we are mainly interested in rhythmical patterns in the sound of samba groove, meaning musical events in time. According to previous research on the groove of samba, the temporal positions of the 16th-notes are of particular interest. Since all the 16th-notes are played, and consequently present in the sound, their temporal positions could be estimated using the onset detection function in the MIR Toolbox (Lartillot & Toiviainen 2007). The onsets calculated from the audio signal may not represent the 16th-notes' perceptual attack points, but since we primarily are interested in the rhythmical pattern, and not necessarily the exact position of the notes, the main concern is that all the onsets are measured in the same manner.

A sound analysis based the second solo recording of the percussionist (I b.) was carried out (see Table 1). The onset data confirmed the above-mentioned systematic micro timing at 16th-note level, that is, the 3rd and the 4th 16th-notes are ahead of their quantized position (Figure 3).

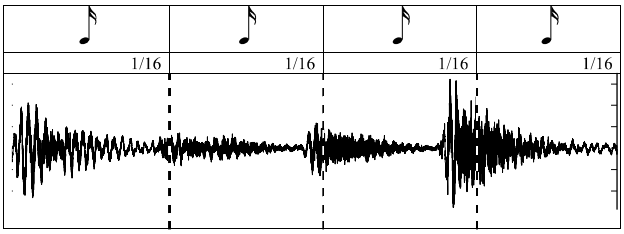


Figure 3: Illustration of the audio waveform and metrical structure of samba at the 16th-note level. The 3rd and the 4th 16th-notes are ahead of their corresponding quantized position.

In accordance with Kvifte's (2004) *pattern concept*, we wanted to investigate the duration pattern of the samba groove. The durations of the 16th-notes were calculated measuring the time from the estimated onset of one 16th-note to the estimated onset of the next 16th-note. The durations of the 16th-notes were measured in seconds with four decimals. Subsequently, the calculated 16th-note durations were converted to percent, according to their percentage of the bar. The duration of 16th-notes in four bars in the beginning of the recording (bar 1 – 4) and four bars in the middle of the recording (bar 13 – 16) were calculated, that is a total of 64 16th-note durations. The mean durations of the 16th-notes, measured in percent, are presented in Table 2.

Sixteenth notes	Mean duration (in %)	SD
First 16th-note	23.0	3.9
Second 16th-note	23.6	6.3
Third 16th-note	22.8	4.4
Fourth 16th-note	30.7	2.1

Table 2: Mean duration and standard deviations (SD) for the first, second, third and fourth 16th-note in a bar (N=64).

Analysis of variance showed significant differences between the durations ($p<0.001$), and Bonferroni corrected post-hoc tests showed significant differences between the duration of the 4th 16th-note and the 1st, 2nd and 3rd (all $p<0.001$). However, no significant differences between the duration of the 1st, 2nd, 3rd 16th-notes were found (all $p>0.99$). A boxplot of the 16th-note durations (Figure 4), and also the standard deviations in Table 2, indicate that not only do the 4th 16th-notes in a beat seem to be of longer duration that the 1st, 2nd, 3rd 16th-notes, but they also seem to be spread out over a smaller range of values than the others. This supports the idea that the long 4th 16th-note in a beat plays a significant role in the groove of samba.

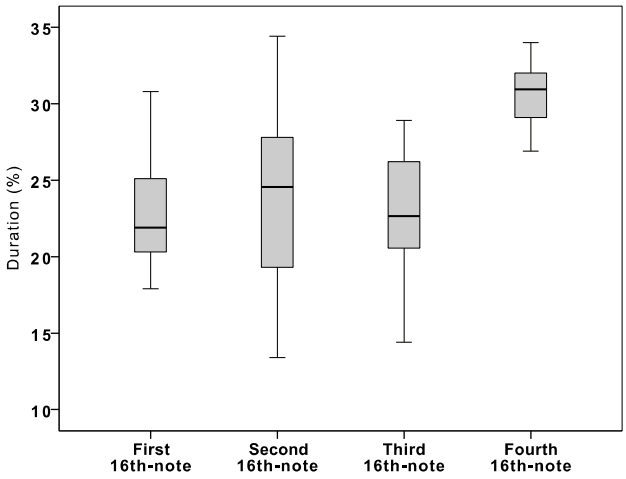


Figure 4: Boxplot showing the distribution of the duration of the 16th-notes in 8 bars. The 4th 16th-note in a beat has longer duration than the 1st, 2nd and 3rd 16th-notes. In addition, the 4th 16th-note durations seem to be spread out over a smaller range of values than the others.

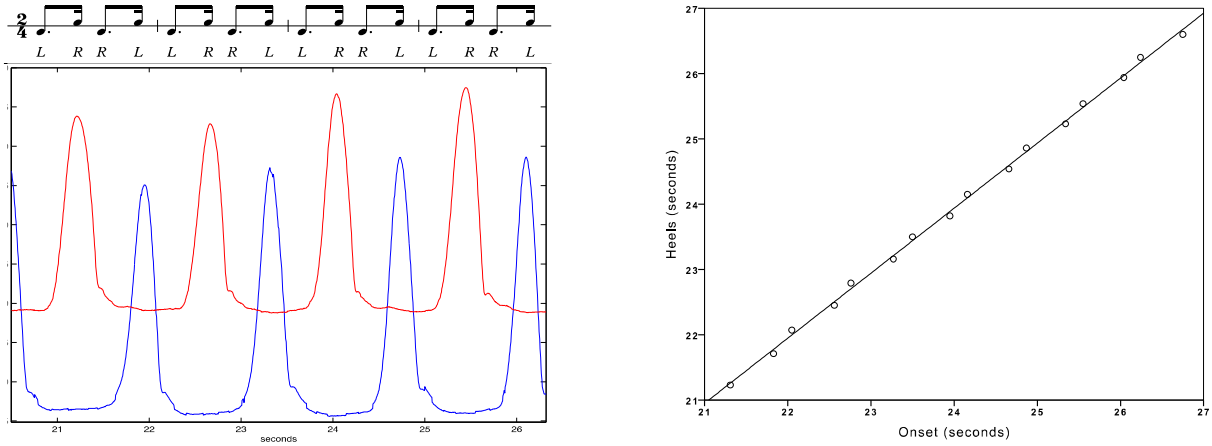
3.2. Motion

The percussionist's feet movements

While playing, the percussionist moves his heels up and down. We wanted to investigate whether the percussionist's heel movements corresponded to the beat level in the sound, and consequently may be interpreted as the percussionist's perceptual pulse ("inner pulsation"). By plotting the heels' vertical position over time we got a visualization of the heels up/down movements. The points where the heels met the floor seems to be in synchrony with the 1st 16th-note in a beat (pulse level) in the sound of samba. Since the lifting of the heels also seems to be periodic, we measured the vertices (local maximum) of the movements of the heels. Comparing the measured vertices of the heel movements with the sound analysis revealed that the lifting of the heels seems to be in synchrony with the 4th 16th-note in a beat (Figure 5a and 5b). Hence, motion analysis of the percussionist's foot movements revealed a motion pattern, not only in correspondence with every 1st 16th-note in a beat, indicating a perceptual pulse level, but also with every 4th 16th-note in a beat, supporting the hypothesis of the essential role of a long 4th 16th-note in a beat in samba.

The dancer's feet movements

The feet movements in samba dance are repetitive and periodic, and we wanted to investigate whether the dancer's feet movements correspond to rhythmical pattern in the sound. Motion data from the second recording of the dancer (II b.) were analyzed (see Table 1). Using the MoCap Toolbox, we made a plot of the vertical movements of the dancer's feet. Next, we manually measured the temporal points where the feet hit the floor, and compared them with the onset data from the sound analysis. The dancer's vertical feet movements seem to be in synchrony with the three first 16th-notes in a beat (Figure 6).



(a) (b)

Figure 5: (a) Plot of the vertical movements of the percussionist's heels, left heel = L (blue) and right heel = R (red). The percussionist's heel movements seem to be in synchrony with both the 1st (pulse level) and the 4th 16th-note in a beat. (b) A scatterplot showing the relation between the onsets (sound) and the percussionist's vertical heel motion.

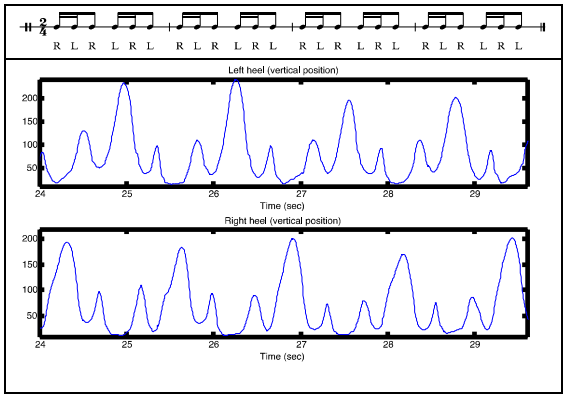


Figure 6: A visualization of the vertical movements of the dancer's heels. The dancer's heel movements seem to be in synchrony with the first three 16th-notes in a beat.

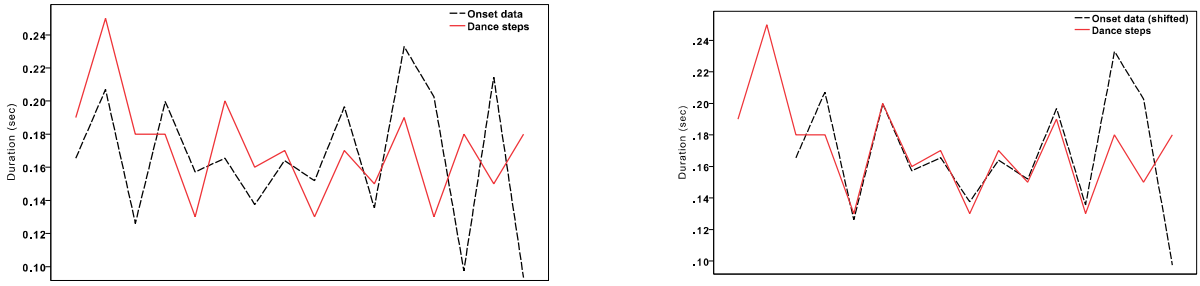
The sound analysis showed that the durations of the first three 16th-notes in the sound seem to fluctuate considerably, and we wanted to investigate how that may influence the dance steps. Equivalent to the beat duration calculations in the sound analysis, we calculated the duration of every step, that is the time between every time the dancer places either of her feet to the floor. The durations of the dancer's steps in bar 13-16 were calculated (N=24). The mean durations of the 16th-notes, measured in percent, are presented in Table 3.

Sixteenth notes	Mean duration (in %)	SD
First 16th-note	22.9	4.5
Second 16th-note	24.1	2.6
Third + Fourth 16th-note	53.0	3.9

Table 3: Mean duration and standard deviations (SD) of the first, second and the sum of the durations of the third and fourth 16th-notes in a bar (N=24).

Analysis of variance showed significant differences between the durations (p<0.001), and Bonferroni corrected post-hoc tests showed significant differences between the duration of the sum of the 3rd and 4th 16th-notes and the 1st and 2nd 16th-notes (all p<0.001). No significant differences between the duration of the 1st and 2nd 16th-notes were found (p>0.99). This corresponds to the results in the audio analysis (see Table 2).

A comparison of the durations of the dancer's steps and the 16th-note duration in the sound, revealed that the dancer seem to copy the 16th-note durations in sound from the previous bar (Figure 7). If this is the case, it may testify to how sensitive humans are to small variations in rhythm, and how these fine-meshed rhythmical patterns can be perceived and reproduced with high accuracy.



(a) (b)

Figure 7: (a) A visualization of 16th-note duration in seconds (y-axis) over time (first two 16th-notes in four bars), for onset data (black) and the dancer's steps (red). (b) The duration of the first two 16th-notes in four bars, onset data (black) shifted one bar. The dancer may copy the durations of the first two 16th-notes from previous bar in the sound.

Establishing the groove

The percussionist always starts the samba groove by playing a start break, generally consisting of two low frequency 8th-notes played on the drum skin with right thumb. Before he starts playing the start break, he makes a smooth rise/fall hand gesture. A plot of the vertical movement of the right hand enabled us to measure the duration of this hand gesture. The hand gesture duration in the two solo

recordings of the percussionist (see Table 1) was measured to 0.75 seconds, which corresponds to the duration of one beat in the samba groove. In other words, the hand gesture together with the start break, which also equals the duration of one beat, has a total duration of a bar (Figure 8). This suggests that the groove is established in the percussionist's body before he starts playing, as an inner simulation of the groove, and that this groove experience is expressed through this hand gesture. This indicates that body motion is intrinsically related to the experience of groove. If the experience of groove, including the groove-related body motion, exists before the percussionist starts to play, it supports the idea that body motion is not only a response to musical sound, but that music and body motion are intrinsically related.

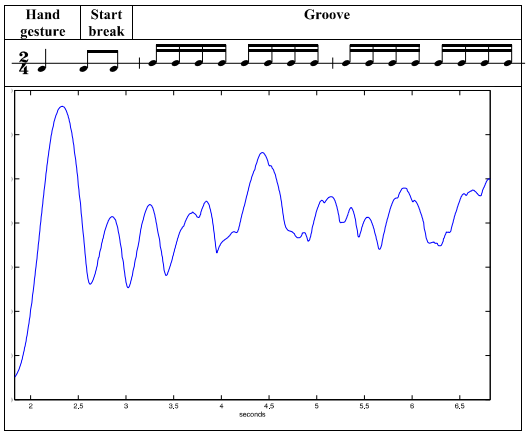


Figure 8: A plot of vertical movement of the percussionist's right hand. The duration of the hand gesture before the start break equals the duration of one beat in the samba groove, suggesting that the groove is established in the percussionist's body before he starts playing.

4. DISCUSSIONS AND FUTURE WORK

We suggest that when investigating microtiming features in samba, body motion may offer a more significant reference level than a more abstract timeline of isochronous points. Consequently, the rhythmical structures of samba were analyzed in terms of durational patterns instead of deviations from quantified durations. The results from our audio analysis of this samba groove, suggest a medium (23%) – medium (23.6%) – medium (22.8%) – long (30.7%) duration pattern on 16th-note level (Table 2). Converting this result to a time line approach, it confirms the so-called anticipation of the 3rd and the 4th 16th-note in a beat, since the combination of the long 4th 16th-note duration and medium 3rd 16th-note duration would cause their onsets being ahead of their corresponding quantized position. However, our result does not correspond to the medium – short – medium – long duration pattern suggested by Gerischer (2006). One explanation may be that Gerischer's analyses are based on samba performances of percussionists from the region of Bahia, while our participants are from outside Rio de Janeiro, i.e. there may very well be local dialects of samba. In addition, our result is based on analysis of only eight bars from one recording of one performer, so the difference may also be due to personal expressive timing. To investigate whether this pattern may be a feature of samba, a larger number of performances need to be investigated.

Our results support the hypothesis from previous research that the 4th 16th-note seems to play a prominent role in samba groove. This hypothesis is also supported by our analysis of the percussionist's movement of the heels, revealing a motion pattern in synchrony with the 1st and the 4th 16-notes in a beat. To investigate the medium – medium – medium – long duration pattern, found in the sound, in relation to body motion, a motion pattern in synchrony with all the 16th-notes is essential. Since, the percussionist's heel movement is only in synchrony with the 1st and the 4th 16-notes and the dancers feet steps are only in synchrony with the 1st, 2nd and 3rd 16th-notes, it does not offer a complete picture of an "inner pulsation". Samba dance is often described as complex full-body movement (Browning 1995), and Mariani (in Naveda 2011) points out that the pulsating rhythm of samba dance originates in the torso. This suggests that the pulsation of samba groove may be integrated in the movement of the hip and torso in samba dancing. In future work this will be investigated further.

In this paper we have analyzed short excerpts of the groove, focusing on rhythmical patterns as musical events in time. In further analysis, we are interested in investigating correspondences between fluctuations in amplitude in the sound and acceleration in the body motion. To get a better understanding of the correspondences between music and body motion in samba, in future work more samba performance recordings, including both musicians and dancers, would be included.

5. CONCLUSIONS

The results of this study point out the importance of including motion data in addition to audio data when analyzing rhythmical structures in samba. Studying rhythmical structures in samba based on recordings of performances, gathering both sound and motion data, allow us to investigate microtiming features in relation to body motion instead of to a constructed time line with isochronous time marks. Our audio analysis confirms systematic microtiming on 16th-level in samba music, and also a synchronized systematic microtiming pattern was discovered in the percussionist's heel tapping and the dancer's steps. These observations support the view that the systematic

microtiming of 16th-notes in samba performance is not a deviation from an underlying perceived pulse with isochronous subdivisions, but constitutes an essential feature of samba.

6. ACKNOWLEDGEMENT

The authors wish to thank Kristian Nymoen for technical assistance during the motion capture recordings, Kathrine Frey Frøslie for helping with the statistical analyses and the percussionist Célo de Carvalo and the dancer Lidia Pinheiro for participating in the study.

7. REFERENCES

Baily, J. (1985). Music Structure and Human Movement. In P. Howell, Ian Cross and Robert West (Ed.), Musical Structure and Cognition *(pp. 237-258). London: Academic press.*

Bengtsson, I. (1987). Notation, motion and perception: Some aspects of musical rhythm. In A. Gabrielsson (Ed.), Action and perception in rhythm and music *(Vol. 55, pp. 69-80). Stockholm: The Royal Swedish Academy of Music.*

Blom, J.P. (1981). Dansen i Hardingfelemusikken. In J. P. Blom, Nyhus, S. & Sevåg,R. (Ed.), Norsk Folkemusikk *(Vol. VII, pp. 298-304). Oslo-Bergen-Tromsø: Universitetsforlaget.*

Blom, J.P. & Kvifte, T. (1986). On the problem of Inferential Ambivalence in Musical Meter. Ethnomusicology, Vol. 30*(No.3 (Autumn)), 491-517.*

Browning, B. (1995). Samba: Resistance in motion. *Bloomington and Indianapolis: Indiana University Press.*

Burger, B. & Toiviainen, P. (2013). MoCap Toolbox – A Matlab toolbox for computational analysis of movement data. Proceedings of the 10th Sound and Music Computing Conference, (SMC), *172-178.*

Carvalho, J.J. (1999). Afro-Brazilian Music and Ritual. Part 1. From Traditional Genres to the Beginnings of Samba*. Série Antropologia 256, Universidade de Brasilia. Lecture presented in the Graduate Seminar in Ethnomusicology at the University of Wisconsin-Madison.*

Gabrielsson, A. (1982). Perception and performance of musical rhythm. In M. Clynes (Ed.), Music, Mind, and Brain. The Neuropsychology of Music *(pp. pp. 159-169). New York: Plenum Press.*

Gerischer, C. (2006). O suingue baiano: Rhythmic feeling and microrhythmic phenomena in Brazilian percussion. Ethnomusicology, 50*(1), 99-119.*

Gouyon, F. (2007). Microtiming in "Samba de Roda" —Preliminary experiments with polyphonic audio. SBCM 2007 Proceedings, *197-203.*

Grau, A. (1983). Sing a Dance. Dance a Song: The Relationship between Two Types of Formalised Movements and Music among the Tiwi of Melville and Bathurst Islands, North Australia. Dance Research: The Journal of the Society for Dance Research, Vol. 1*(No.2 , Autumn), 32-44.*

Iyer, V. (2002). Embodied Mind, Situated Cognition, and Expressive Microtiming in African-American Music. Music Perception, 19*(3), 387-414*

Keil, C. (1987). Participatory Discrepancies and the Power of Music. Cultural Anthropology, 2*(3), 275-283.*

Kubik, G. (1990). Drum Patterns in the "Batuque" of Benedito Caxias. Latin American Music Review / Revista de Música Latinoamericana, 11*(2), pp. 115-181.*

Kvifte, T. (2004). Description of grooves and syntax/process dialectics. Studia musicologica norvegica, 30, *54-77.*

Kvifte, T. (2007). Categories and Timing: On the perception of Meter. Ethnomusicology, Vol. 51*(No. 1), 64-84.*

Lartillot, O. & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction From Audio. International Conference on Digital Audio Effects.

Mariani, M.E. (1998). African Influences in Brazilian Dance. In K. W. Asante (Ed.), African Dance *(pp. 79-97). Eritrea: Africa World Press.*

Naveda, L. (2011). Gesture in Samba: A cross-modal analysis of dance and music from the Afro-Brazilian culture. (Ph.D. thesis), Faculty of Arts and Philosophy, Ghent University, Belgium, Ghent, Belgium.

Roy, P., Pached F. and Krakowski S. (2007). Analytical features for the classification of percussive sounds: The case of the pandeiro. Proceedings of the 10th International Conference on Digital Audio Effects (DAFx '07), pp. 213-220.

Shove, P. & Repp, B.H. (1995). Musical motion and performance: theoretical and empirical perspectives. In J. Rink (Ed.), The Practice of performance: studies in musical interpretation (pp. 55-83). Cambridge: Cambridge University Press.

CAPOEIRA INTERACTION AS A MODEL OF EXPECTATION FORMULATION AND VIOLATION IN REAL-TIME IMPROVISED PERFORMANCE

Megha Makam[1], Blair Kaneshiro[2], Jonathan Berger[2]
[1]Department of Biology, Stanford University, USA, [2]Center for Computer Research in Music and Acoustics, Stanford University, USA
Time: 17:00

ABSTRACT

Capoeira is an Afro-Brazilian martial art uniquely driven by music. Paired bouts (called "games") are characterized by continuous stepping, kicking, and sweeping movements performed to live music. Capoeira movements can be executed with the intention to maintain a cooperative, fluid interaction with the opponent, or as overt attacks (sweeps, strikes, or takedowns), and thus integrate patterns that generate high expectation with temporal surprise. We interpret the predictive aspect of the capoeira bout as a useful model for studying expectation formulation, realization, and violation in an improvisatory context analogous to musical engagement by music

improvisers. In addition to the "performer interactions" of the paired opponents, we consider the connection between music and movement in the context of capoeira.

1. INTRODUCTION

One of the most dynamic manifestations of human physical ability is demonstrated in the Afro-Brazilian martial art of *capoeira* (ko'pwera). Capoeira is a "blurred genre", combining elements of dance, folklore, martial arts, sport, ritual, and training for unarmed fighting (Downey, 2002). Originating in sixteenth-century Brazil as a consequence of the intercontinental slave trade between the South American and African continents, capoeira has evolved in various populations of Brazilian society (Taylor, 2005). Today, the rapidly growing community of capoeira practitioners, or *capoeiristas*, forms an international network of people who cultivate and spread Afro-Brazilian culture through practice of this art.

We consider capoeira as an ecologically relevant real-time model of improvisation, which can be used to study how expectation is formulated, then realized or violated. Furthermore, as a movement art, it is amenable to visual recordings and thus advantageous for studying gross body movements. Capoeira is unique in that it is performed and practiced in a specific, tradition-based musical context, yet it remains unclear how exactly the movements correlate with musical events. The main goal between two capoeira players is to interact fluently, as in a conversation, while maintaining an appropriate level of surprising moments (often in the form of attacks like sweeps, strikes, or takedowns). These interactions inspired the following questions: How is this ideal balance achieved between players? What types of movements help construct surprising moments? When, in time, do these moments occur, and how is the physical space utilized to this end? Does music play a role in the construction of surprising moments?

Jazz improvisation is currently regarded as a useful setting for studying creativity and interpersonal interactions. While we can design theoretical models to predict improvisatory interactions within a jazz ensemble (Hudak & Berger, 1995) or study jazz improvisation in highly controlled brain-imaging experiments (Limb & Braun, 2008), there are immediate questions about the ecological validity of such contexts (McPherson & Limb, 2013). We propose that interactions between capoeira players as they improvise together can be quite analogous to improvisers in a jazz musical context. Furthermore, capoeira interactions are likely translatable to other musical contexts and even broader creative behaviors.

This paper will follow three stages of discussion to frame capoeira performance as a physical manifestation of musical expectation formation and violation. First, we introduce readers to the fundamentals of capoeira, which includes an overview of principal movements, musical elements, and their synthesis as *malícia*, the over-arching philosophy that characterizes the mindset of a capoeirista. Next, we explore the improvisatory features of capoeira, characterizing successful interactions and drawing analogies to jazz improvisation. Finally, we discuss the potential of capoeira as a novel model for studying performance interaction and manipulation of expectation in music.

2. ANATOMY OF CAPOEIRA

A typical capoeira event, called a *roda* (ho – da), comprises a circle of people singing in Portuguese and clapping as some members play percussion instruments, while two people in the center physically interact with fluid, combative, yet sportive earnestness. Though capoeira may be interpreted in many ways, it is ultimately play between people, containing the seeds of all forms of interaction (Downey, 2008). Understanding the process of capoeirista skill acquisition affords insight into the elements of the performance.

Capoeira is learned primarily by imitation in classes taught by either a *mestre* (master instructor) or an advanced student. The learning process is facilitated by a tight link between the perception of others' actions and one's own sense of self. Although movements are taught through mimicry, it is expected that the execution of the movements in a game context will be improvised. Moreover, so many factors (fatigue, speed, position relative to the opponent, etc.) influence the external condition that every time a specific "movement" is used, it will never be identical to one completed previously. There also exists a large link between the performance of movements and the musical environment – often the pace of the interaction between players follows the musicians' tempo, though novices often move more rapidly than expected with the music (Downey, 2005).

2.1. Movements

Movements described here introduce only the very basic elements of the physical aspects of capoeira, and provide at most a starting point for further inquiry. Furthermore, different styles and variations on the same basic positions mean that no instructor has the same form, though essential movements would be identifiable.

The very first movement learned by students is the ginga (jin – ga), which is the distinct swaying motion that characterizes capoeira.[1] It is the "point of departure for all future acquisitions…'fundamental' position of the Capoeirista…the key to his agility and evasiveness" or the "ultimate signature of the expert" (Downey, 2005). The ginga is the glue connecting all other movements, facilitating fluidity and allowing the player constant momentum for defenses or attacks, all while concealing one's intentions. Moves are generally categorized as either a defense or attack, and a few will be described here.

Defensive movements. As avoiding injury is a priority in any capoeira interaction, one learns to escape harm using defensive movements while also keeping constant view of the opponent. The negativa (neh – ga – chee – vuh) is a quick squat to the floor with one leg bent and the other outstretched. This is an efficient way to avoid an opponent's attack to the upper body. A basic cartwheel, called aú, is also used to escape attacks, and opens many possibilities for subsequent movements while translocating to an unexpected location. The

1 An example of the ginga being performed can be viewed on Wikipedia:
http://upload.wikimedia.org/wikipedia/commons/9/9a/Ginga_de_dos.gif

handstand, or bananeira, can be used to rest one's feet, while also maintaining a defensive position with feet ready to kick if necessary while maintaining view of the opponent.

Attack movements. One of the most characteristic offensive movements is the foot sweep, called the rasteira (ha – shtey – ra). The effectiveness of this movement strongly relies upon anticipating of the opponent's position and correctly timing the release of the sweep. There are many variations of kicks that can be used, and will not be described here, though they all can exploit the opponent's various vulnerable positions with proper timing. Another common attack is the head-butt, or cabeçada (ka – bay – sah – dah).

2.2. Music

In order to fully participate in the practice of capoeira, students also learn to play musical instruments and sing songs. The capoeira audience, also called the roda, stands around the two players as they engage in playful combat. The capoeira instrumental section, the *batería*, comprises several instruments that are played by capoeiristas as part of the roda. There are three sizes of *berimbau*, a monochord with a gourd as resonator, which is moved away and towards the musician's stomach as he strikes the steel wire, which changes the pitch and volume. The three different berimbaus played together often fulfill different roles in cuing particular events in the game through specific rhythmic phrases, particularly the lowest pitched berimbau. Each berimbau is used to play complementary rhythmic patterns that interlock and involve much improvisation. Most of the time, one of the berimbau players also leads the group in singing the songs. The batería additionally includes the *atabaque* (ah – tah – bah – key) (a tall hand drum), *pandeiro* (small frame drum with jingles), *agogô* (double bell), and *reco-reco* (heh – ku – heh – ku) (scraper). The musicians and capoeira players constantly exchange roles so that all participants usually fulfill all the roles of either playing capoeira or music.

Typical capoeira songs are sung in a set format, where the roda usually begins with a *ladainha* (la – da – EE – nya), a solo song that often has spiritual meaning for the group accompanied only by berimbaus, followed by the *chula*, which invites players in the first call-and-response verses. The percussion enters in the last type of song, the *corrido*, which allows the two players in the center to start their game. The call-and-response of the songs also invites all participants to sing and clap their hands during the performance, which encourages the two players in the middle (Talmon-Chvaicer, 2008).

2.3. Malícia in Game Dynamics

Malícia in capoeira is cunning – a savvy that can manifest as humor, technical virtuosity, deceptiveness, ability to anticipate another's actions, superior command of space, and a sense of dramatic or malicious opportunism. Capoeiristas embody this attitude by anticipating an opponent's rhythms in order to disrupt their flow, a goal that differs from many other dance contexts. To an unfamiliar observer, it may seem that advanced players "groove" together, but this interaction is one in which anticipation between players follows so closely in time that they appear to move with the same flow and with the music. Some masters of capoeira caution players to avoid depending too much on eyesight, as they can be deceived with feints and visual distractions by more experienced players; they advise that sound can reveal opportunities, where one may "listen" for the right moment to attack (Downey, 2002). This advice demonstrates the contradictory relationship that capoeiristas may have with music; to keep a balance between fluidity and avoiding becoming trapped, while also initiating attacks, requires a constant monitoring of both auditory and visual streams.

As previously mentioned, the dynamic stance of the ginga exemplifies this ambiguous relationship of a player's movements with music – the footfalls of the ginga do not fall symmetrically with the pulse of the music, allowing the player to initiate subsequent movements from any position within the stance. This is an essential and unique characteristic of capoeira that enables players to conceal their movements or mislead the opponent to expect something else. This underlying attitude of opportunism and anticipation shapes capoeira players' relationship with the musical environment.

3. TIMING, PREDICTION, AND IMPROVISATION

Capoeira is a unique, movement-based art form that fits in a category of its own, separate from both dance (where focus is on feeling and evaluating one's movement, proprioception, kinesthesia, and visual image as aesthetic style) and martial arts (where the artist attends to opponents as objects, acting through the body to attack and defend oneself). In capoeira, there is a state of consciousness either divided between modes of awareness or in which one moves rapidly between modes (Lewis, 1995). This consciousness is also something developed through practice, as there is an implicit cultural expectation that capoeira players look good even when forced by chance events to improvise movements in unexpected ways in response to opponents, musicians, and the audience. These improvised movements rely on the ability of players to quickly and accurately anticipate events in their environment.

3.1. Grounding in Capoeira Interactions

The interactions between players during capoeira performance can be described using the same terms that describe interactions between jazz musicians while they improvise. The process of grounding that occurs in both of those contexts, or the continual monitoring of shared understanding, is fundamental to a successful capoeira interaction. Timing is a main feature of grounding in these collaborative social interactions, and it facilitates moments that can be anticipated and used to support coordinated behaviors. While musicians signify and display their understanding of each other's expressions, they manage the common ground and create a new common ground upon which to act (Gratier, 2008). Similarly, capoeira players communicate through facial and bodily gestures and overt movements to do this.

When capoeira players of varying experience are asked what they consider desirable attributes of a capoeira game, most will report the aspect of flow and conversation that they experience with the other player (personal interviews). This is understandable, as the emotional and aesthetic value of anticipation is known; the release of tension is emotionally gratifying and intrinsically meaningful

(Huron, 2006). Conversations – whether verbal, musical, or physical – utilize grounding devices that can help individuals achieve the desired interaction, increasing fluency and rhythmic flow.

Acknowledging and displaying anticipation, repetition or mirroring, and synchronization all facilitate this flow experience. Capoeira players will "check" each other, showing that they have predicted their opponent's position. Often, less experienced players will mirror their opponents to achieve the experience of connecting – since it is safer than initiating attacks – without the ability to predict the more experienced player's position. The well-timed patterns of repetition and variation involve various degrees of predictability, and anticipation relies on awareness of this as well as implicit timing ability.

Meanwhile, similar to jazz improvisers who integrate personal expressivity while respecting the overall principles of a given piece and its style, capoeira players also learn to navigate the framework of their art as they gain experience. Additionally, "being in synch" is fundamentally necessary but not sufficient for improvisation that will be aesthetically or pragmatically successful; this fact applies to both jazz and capoeira improvisation. Due to the ambiguous and flexible nature of capoeira interactions, practitioners prepare during training to use these grounding devices (though not explicitly). The games can range from artistic cooperation to violent conflict, and training must prepare a player to synchronize with both the opponent and the music while also being prepared to break synchrony in order to avoid being struck, or to initiate an attack before the opponent realizes it has occurred. Experienced capoeira players develop the ability to change direction in the middle of a movement, integrating feints and misleading their opponent, all of which is done by establishing a conversational flow using the aforementioned grounding devices. It is worth noting that when two capoeira mestres play each other, each has such a well-honed ability to anticipate the other's movements that truly intentional attacks are often subtler and more difficult to detect by an untrained observer.

3.2 Setting Up a Surprise

In terms that have previously been used to describe construction of surprise in musical compositions, expectation violations operate along two dimensions of the musical event ("what") and when it occurs ("when"), while the dimension of "where" relates to the expectation of sequential ordering of particular events (for example, dominants resolve to tonics). We can also describe these events in capoeira interactions, where:

a. "What" is the movement, (sweep, kick, or takedown).

b. "When" indicates the timing of the movement and its place within a continuous sequence of movements.

c. "Where" is the movement's spatial coordinates (for example "low" or "high" kick, etc.).

As we proceed in our research, we will explore the expectation-violation model in capoeira as analogous to music. A large corpus of capoeira attacks in real-world performances is available for analysis. By first characterizing the ginga and other flow-based movements in relation to the accompanying music, we hope to reach a detailed understanding of the contextual violations that successful capoeira attacks embody.

Overall, the game's intention is to exploit opportunities that the opponent's movements provide and avoid being caught oneself. The interactions are necessarily improvisatory and fluid, and as Charles Keil posits, an improvising group needs a framework in time, which music provides (Keil, 1966).

3.3 The Mysterious Role of Music

Many players report that the music during the roda gives them energy to play, as the lyrics of the songs often encourage them, advise them, or reflect events occurring between the players. The berimbaus' complex rhythmic patterns over the steady and regular beat of the drum provide a temporal structure within which the players move. Interestingly, as the berimbau is rarely heard outside of the capoeira environment, it can become strongly associated with the act of playing capoeira and the arousal state that is required to perform it. Players also report that a disturbance in the musical stream can bring their physical interaction to a halt.

Furthermore, we may understand the tactical departures from synchronicity that occur in capoeira, also called "participatory discrepancies", where players create tension and test each other's skill while also interacting with the musical environment (Keil, 1987). This interaction is what characterizes the relationship between players in capoeira, embodying the concept of malícia.

Ethnographic and anecdotal descriptions of performers' experience indicate that there exists a unique, indirect relationship between music and movement during capoeira performance. While the primary objective for two performers interacting is to maintain a fluid conversation while managing an ideal level of surprising events, the exact influence of the musical context remains elusive. Our preliminary research suggests that the movements executed during game play are the emergent result of numerous complex interactions that occur in the presence of music, though they need not be directly related to the music.

4. DISCUSSION

We have introduced the craft of capoeira, described a few of its fundamental movements, and characterized the musical setting in which capoeira games are performed. The temporal framework from the music provides the common ground necessary for successful improvisation, and it is the setting of musical expectations using this framework that allows for the deviations that constitute successful capoeira attacks. However, though the music plays a large role in providing this framework, some practitioners report that during particularly high-energy or aggressive interactions, they no longer are conscious of the music, rather allocating more attention toward anticipating their opponent's movements (personal interviews). Meanwhile, others hold that "playing capoeira without music amounts to doing pushups (deMoor). These seemingly contradictory reports highlight the very reason capoeira should appeal to cognitive scientists.

The interaction between the capoeira players and their musical environment is an untapped experimental context that may help us to better understand the dynamical relationship between music and improvised movement in an ecologically valid setting. The cultural acquisition of bodily patterns of responsiveness and attentiveness are instilled through habituation and training, providing a potentially useful model to bridge the anthropological understanding of embodiment with a more biological and neuroscientific one (Csordas, 1993). While listening to capoeira music, practitioners feel the swaying movements in memory through the training, as outward movement or inward quickening, or readiness to move (Downey, 2002). However, players must reallocate their attention from the music to other elements in the environment in order to protect themselves while maintaining a fluid interaction, and this may not necessarily be synchronized with the music. To position the study of this art among existing studies on movement, dance, and instrumental performance would broaden the discussion and generalize results in a new way.

Our hope is that the analysis presented in this paper provides a foundation for future capoeira-based research. Capoeira has the potential to inform insights about a variety of fields, ranging from ethnographic analyses to movement studies to performance practice, and may be applied to areas as broad as video game design and military training. Currently underway is an ethnographic study specifically investigating how capoeira players from a variety of schools and pedagogical styles relate to their musical environment and their reported experiences of how it influences their movements in the roda. Furthermore, insights gained from first-hand reports of capoeira practitioners will inform our experimental design of empirical studies aiming to investigate, for example, whether the relationship between ginga footfalls and musical pulse can be quantified or characterized.

5. REFERENCES

Csordas, T. J. (1993). Somatic Modes of Attention. Cultural Anthropology, 8(2), 135-156.

deMoor, M. (n.d.). Jogo de Capoira Angola. Retrieved May 31, 2014, from www.wesleyan.edu: http://www.wesleyan.edu/wsa/capoeira/articlejogo.html

Downey, G. (2002). Listening to Capoeira: Phenomenology, Embodiment, and the Materiality of Music. Ethnomusicology, 46(3), 487-509.

Downey, G. (2005). Learning Capoeira: Lessons in Cunning from an Afro-Brazilian Art. Oxford: Oxford University Press.

Downey, G. (2008). Scaffolding imitation in capoeira: physical education and enculturation in an Afro-Brazilian art. American Anthropologist, 110(2), 204-213.

Gratier, M. (2008). Grounding in musical interaction: evidence from jazz performances. Musicae Scientiae, 12, 71-110.

Hudak, P., & Berger, J. (1995). A Model of Performance, Interaction, and Improvisation. Proc. Int. Computer Music Conference. International Computer Music Association.

Huron, D. (2006). Sweet Anticipation: Music and the psychology of expectation. Cambridge, Massachusetts: MIT Press.

Keil, C. (1966). Motion and feeling through music. The Journal of Aesthetics and Art Criticism, 24(3), 337-349.

Keil, C. (1987). Participatory Discrepancies and the Power of Music. Cultural Anthropology, 2(3), 275-283.

Lewis, J. L. (1995). Genre and embodiment: from Brazilian capoeira to the ethnology of human movement. Cultural Anthropology, 10(2), 221-243.

Limb, C. J., & Braun, A. R. (2008). Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation. PLoS ONE, 3(2), e1679.

McPherson, M., & Limb, C. J. (2013). Difficulties in the neuroscience of creativity: jazz improvisation and the scientific method. Ann. N.Y. Acad. Sci.(1303), 80-83.

Talmon-Chvaicer, M. (2008). The Hidden History of Capoeira: A collision of cultures in the Brazilian battle dance. Austin: University of Texas Press.

Taylor, G. (2005). Capoeira: The Jogo de Angola from Luanda to Cyberspace (Vol. 1). Berkeley, CA: North Atlantic Books.

THE RELATIONSHIP BETWEEN TEMPORAL ANTICIPATION AND ADAPTATION DURING SENSORIMOTOR SYNCHRONISATION.

Peta Mills[1], Peter Keller[1]
[1]The MARCS Instituite, University of Western Sydney, Australia
Time: 17:30

ABSTRACT

Experienced music and dance ensembles can coordinate their movements with extreme temporal precision and accuracy, and yet remain flexible during constantly changing conditions. This sensorimotor synchronization requires individuals to continuously anticipate and adapt to each other's action timing. Individuals differ in their ability to both anticipate and adapt, however little is understood about the relationship between these skills. The present study used paced finger tapping tasks to examine the relationship between anticipatory skill and adaptive (error correction) processes. In addition, the contribution of anticipatory and adaptive mechanisms, to individuals' synchronization precision and accuracy was investigated. Adaptive ability was estimated by the degree of temporal error correction that participants (N=52) engaged in when synchronizing with a 'virtual partner', that is, an auditory pacing signal that modulated its timing based on the participant's performance. Anticipation was measured by calculating a prediction index that reflected the degree to which participant's inter-tap intervals led or lagged behind inter-onset intervals in tempo-changing sequences. A correlational analysis revealed a significant positive relationship between prediction/tracking indices and error correction estimates, suggesting that temporal anticipation and adaptation interact to facilitate synchronization performance. Hierarchical regression analyses revealed that adaptation

was the best predictor of synchronization accuracy, whereas both adaptation and anticipation predicted synchronization precision. Together these results demonstrate a relationship between anticipatory and adaptive mechanisms and indicate that individual differences in these two abilities are predictive of synchronization performance.

1. INTRODUCTION

Humans have the ability to coordinate movement with others with little apparent cognitive effort, for example, shaking the hand of another or clapping hands in synchrony. This is exemplified by experienced music and dance ensembles who coordinate their movements with extreme temporal precision and accuracy, and yet remain flexible during constantly changing conditions. Such rhythmic interpersonal coordination requires the combination of many perceptual, motor, cognitive and social processes (Knoblich, Butterfill, & Sebanz, 2011; Phillips-Silver & Keller, 2012). Two core cognitive- motor skills that musicians employ to successfully synchronize are anticipatory and adaptive skills (Keller, 2008; 2014). Musicians must continuously make predictions about upcoming sounds and movements while also adapting their movements in response to previous timing deviations. Prior research (e.g. Fairhurst, Janata, & Keller, 2012; Pecenka & Keller, 2011) has identified that individuals differ in their ability to both anticipate and adapt, however little is understood about the relationship and interaction between these skills and how they work together to enable rhythmic interpersonal coordination.

Anticipatory and adaptive skills serve complementary functions in supporting synchronized movement (van der Steen & Keller, 2013). Temporal anticipation is based on prospective information whereas adaptive timing relies on retrospective information. The former is predictive while the latter is reactive. Temporal anticipation allows a performer to predict the timing of others' upcoming movements in order to plan and coordinate their own movement. On the other hand, adaptive timing engages error correction, where performers respond to timing variations (intended expressive timing deviations or unintentional errors) by making adjustments to their own movement timing. Both anticipation and adaptation contribute to the quality of timing accuracy and precision displayed by ensemble performers (Keller, 2008; Pecenka & Keller, 2011; Repp & Keller, 2008), however the relative contribution and interaction between the two skills is yet to be determined.

Temporal anticipation and adaptive timing have previously been studied extensively (but separately) within the context of sensorimotor synchronization tasks. These sensorimotor synchronization tasks require participants to tap in time with auditory pacing sequences (for reviews see Repp, 2005; Repp & Su, 2013). This simplified task allows controlled investigation of the mechanisms involved in interpersonal joint action (Repp, 2005; Repp & Su, 2013). Synchronization is generally assessed in terms of accuracy and precision. Accuracy is how close a person's taps are in relation to the tones, and is assessed by measuring synchronization errors or asynchronies (defined as the temporal difference between the onset of a tone and a tap). Precision refers to how stable or consistent a person's tap timing is, and is assessed in terms of variability (how much timing fluctuates between successive taps).

1.1. Adaptive Timing (Error Correction).

Adaptive timing allows musicians to maintain interpersonal synchrony in the face of each other's intentional and unintentional timing deviations. Adaptive timing is mutually employed by multiple musicians, each responding to asynchronies by the adjustment of their own subsequent actions (Goebl & Palmer, 2009). Adaptive timing is supported by error correction mechanisms that enable internal timekeepers—oscillations of neural populations in an individual's brain—to remain entrained with a sequence of pacing events. (Repp, 2001, 2011; Repp & Keller, 2008; Vorberg & Schulze, 2002). There are two types of error correction processes, period and phase correction, each an independent process that functions to reduce asynchrony in timing (Mates, 1994; Vorberg & Schulze, 2002). Phase correction is an automatic process that corrects deviations in timing continuously, adjusting the timing of each movement based on the previous asynchrony while leaving the period (the interval of time between successive events) of the internal timekeeper unchanged. This process occurs at a millisecond timescale even without conscious awareness of asynchrony. Period correction on the other hand is an intentional adjustment of the internal timekeeper, resulting in a change of movement tempo (Repp, 2005). This form of error correction requires conscious perception of a tempo change in the pacing sequence. (Keller, 2008, 2012; Repp, 2001, 2005; Repp & Keller, 2004).

Error correction, in particular, phase correction has been investigated using a 'virtual partner' (e.g. Fairhurst, Janata, & Keller, 2012), an adaptive metronome that enables the computer to engage in controlled simulation of error correction during a tapping synchronization task and thus 'interact' with a participant (Repp & Keller, 2008). To simulate error correction, the computer uses a mathematical algorithm to respond to an asynchronous tap by altering the timing of the subsequent tone. For example during phase correction, if a participant were to tap slightly too early, the computer will respond by adjusting the timing of the following tone to sound earlier, (see Figure 1). The proportion of the asynchrony compensated for in phase correction is controlled by the parameter 'alpha (α)'.

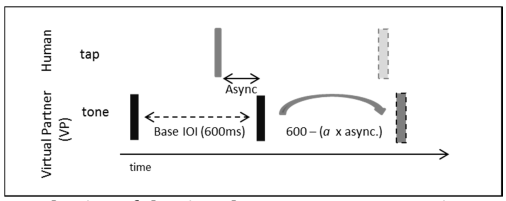


Figure 1: Overview of the adaptive timing mechanism of the virtual partner. Error correction mechanisms alter the timing of the following pacing event by adjusting for a proportion of the previous asynchrony. The timing of the inter-onset interval (IOI) is adjusted by a proportion (*α*) of the asynchrony between the previous pacing event and tap.

By adjusting α, the degree of phase correction employed by the virtual partner can be prescribed and manipulated by an experimenter. By varying the degree to which the computer adjusts the timing, it is possible to make the computer simulate either a cooperative or uncooperative musical partner (Fairhurst et al., 2012; Repp & Keller, 2008). This is analogous to actual human partnering where the skill level of musical partners is variable. Coupling strength is optimal (defined as minimal variability of asynchronies) when both the virtual partner and the participant engage in moderate amounts of phase correction. Both the human participant and the virtual partner each moderately adjust their timing toward each other, minimising the variability of asynchronies. However, if the computer is programmed to be overly adaptive and over-correct the asynchrony, this actually hinders synchronization.

By employing the virtual partner, it is possible to ascertain what degree of assisted synchronization is optimal for each individual. Once this is known, an estimate of each individual's own error correction capacity (the mean proportion of each asynchrony that is corrected) can be obtained (Repp & Keller, 2008). These estimates of individual's error correction (referred to as 'human alpha') are based on the autocorrelation of asynchronies derived from synchronization tasks with the virtual partner at various degrees of adaptivity.

1.2. Temporal Anticipation

Anticipatory mechanisms allow an individual to accurately predict or anticipate others' upcoming movements and sounds. In musical contexts, co-performers predict the timing of subsequent sounds based on regular rhythmic patterns within the music. Anticipatory mechanisms also allow musicians to predict changes in musical timing (e.g. tempo changes), such as those that occur during expressively motivated musical performance (Pecenka & Keller, 2009; Rankin, Large & Fink, 2009). Keller, Knoblich, and Repp (2007) proposed that the anticipation of co-performers' actions is driven by action simulation processes that are supported by internal models. Internal models allow mental simulation of a movement, and the potential outcome of such a movement, to be carried out prior to action execution. In order for this to occur, the central nervous system must have had prior experience of contingencies between efferent neural motor signals and afferent sensory information and their effects on the body and environment (Wolpert, Miall, & Kawato, 1998). It has been suggested that internal models aid movement efficiency by allowing predicted sensory (afferent) feedback to inform the accuracy of a movement, before the arrival of actual sensory feedback (Aschersleban, Stenneken, Cole, & Prinz, 2002; Keller et al., 2007; Wolpert et al, 1998). Internal models thus allow a kind of 'anticipatory error correction', adjusting planned movements in order to correct potential errors before they occur.

Two types of models, forward and inverse, are hypothesized to operate concurrently for not just execution of one's own actions but also for emulating and predicting the actions of others (Keller 2008, 2012; Pacherie, 2008; Wolpert et al., 1998). Through observation of others' actions, forward models allow internal simulation and thus calculation of others' movement outcomes. Inverse models on the other hand enhance synchronization by representing others' intentions and performance goals and using this knowledge to predict what actions will subsequently be produced. It is theorised that the use of forward and inverse models for both self and other enables anticipation of forthcoming actions or sounds, which forms an important basis of anticipatory ability (Keller, 2008). In support of this hypothesis, Stoit et al. (2011) found that deficits in the ability to model the behaviour of a co-actor (e.g. as occurs in autism), impairs performance in joint action tasks.

Action simulation processes supported by internal modelling mechanisms allow for deliberate mental imagery of co- performers' upcoming movements and sounds (Keller & Appel, 2010). It has been proposed that such imagery is based on knowledge of the shared representations and goals of the ensemble (Keller, 2008). The ability to engage in mental simulation and imagery can facilitate coordination. For example, Keller and Appel (2010) found that individual differences in anticipatory auditory imagery predicted the quality of coordination in piano duos. The simulation and imagery required to accurately predict others' movements relies on cognitive resources such as working memory. Pecenka, Engal, and Keller (2013) found increased cognitive load associated with a working memory task led to decreased prediction ability during a simultaneous sensorimotor synchronization task.

Individuals differ markedly in anticipatory ability, with some being very good at predicting tempo changes while others follow or 'track' these changes (Rankin, Large & Fink 2009; Pecenka & Keller, 2011). In an investigation of individual differences in anticipatory ability, Pecenka and Keller (2011) calculated a prediction index as a measure of individual anticipatory ability. This measure was based on a sensorimotor synchronization task with a tempo changing auditory sequence. The cross-correlation between participant's inter-tap intervals (IOIs) and the inter-onset intervals (IOIs) of the tones in the auditory sequence were computed at both lag-0 and lag-1. The lag-0 cross correlation indicates how accurately the participant anticipated or predicted the timing of the subsequent tone, whereas the lag-1 cross correlation assesses to what degree the participant's taps matched the timing interval of the previous tone. A ratio of the lag-0 over the lag-1 provides an index of the degree to which each person was able to anticipate (predict) tempo modulations in a tempo changing sequence as opposed to following (tracking) these changes (Pecenka and Keller, 2011; Rankin et al., 2009; Repp, 2002). The results indicated that individual differences in anticipatory ability were stable across time. Moreover, synchronization performance on a dyadic tapping task was found to be better when two high predicting individuals (those who anticipated tempo changes more than followed tempo changes) were paired together compared to low predicting pairs (those that followed more than predicted tempo changes) or mixed pairs comprising high and low predicting individuals.

1.3. The Interaction Between Anticipation and Adaptation

Until recently, each of the mechanisms that support musical joint action have been investigated separately, so the relationship and interaction between these mechanisms is poorly understood. Van der Steen and Keller (2013) addressed this by introducing the ADaptation and Anticipation model (ADAM), a computational model that implements temporal anticipation and adaptive timing within a single framework. To complement this modelling approach, the current study employs a behavioural sensorimotor synchronization task to empirically investigate both anticipation and adaptation within the one study. The main aim was to examine the relationship between

individual differences in anticipatory skill and adaptive timing. A secondary aim was to assess the contribution of anticipatory and adaptive mechanisms to individuals' synchronization precision and accuracy.

Several possible outcomes were envisaged. First, it is possible that temporal anticipation and adaptive timing may be independent processes. Anticipation is a cognitive process (Pecenka et al., 2013) whereas adaptation, specifically phase correction, occurs automatically without awareness (Repp, 2005), which raises the possibility that the two processes are unrelated because they call upon separate perceptual and motor resources in the brain. On the other hand, the hypothesis that anticipation utilises internal models that engage in anticipatory error correction suggests that these two processes may collaborate within the sensorimotor control system. This relationship between anticipation and adaptation may, however, take one of two forms. The relationship may be negative in the sense that less error correction is required when higher prediction is employed, because asynchronies would be small in the first place. Alternatively the relationship may be positive, with those who are better at predicting tempo changes also displaying higher phase correction. This may be the case because those with higher anticipatory tendencies tend to have overall smaller asynchronies and less error to correct (Pecenka & Keller, 2011), therefore larger proportions of the asynchrony can be corrected without disrupting the stability of movement timing. With regard to the second aim of the present study, it was hypothesised that both anticipation and adaptation would be good predictors of synchronization accuracy and precision, as both mechanisms have been linked to reduced asynchrony and variability (Pecenka & Keller, 2011; Repp & Keller, 2008).

2. METHOD

2.1. Participants

A total of 69 individuals participated in the study (57 female, 12 male; age range = 18 – 50; *M* = 25.81, *SD* = 8.86). This group comprised first-year psychology students who participated in return for course credit and voluntary associates of the MARCS institute at the University of Western Sydney. The majority of participants (65%) reported having no musical experience. Of the 24 participants who reported having some form of musical training, 18 had between one and four years, while 6 reported having 6 or more years of musical training. After excluding participants who recorded insufficient tapping data (see below for more details), 52 participants remained (43 female, 9 male; age range = 18 -50; *M* = 27.02, *SD* = 9.41) of which four reported having >6 years of musical training. All participants had self-reported normal hearing and all provided informed written consent prior to participation.

2.2. Design

The experiment consisted of three synchronized tapping tasks: Isochronous, tempo changing, and adaptive. The isochronous tapping task provided a measure of general synchronization ability, including assessment of both precision and accuracy. The tempo-changing task provided a measure of anticipatory ability through calculation of prediction indices. Finally, the adaptive virtual partner task was employed with varying levels of computer adaptivity to gain an estimate of each participant's adaptive tendency. The isochronous tasks and the adaptive virtual partner tasks were presented at both fast (400ms) and slow (600ms) tempi.

2.3. Materials and Stimuli

Participants tapped on a Roland Handsonic 10 percussion pad connected via an M-Audio MIDI interface to a MacBook Pro laptop running Max/MSP software. The auditory stimuli were sequences of percussion tones presented over Sennheiser HD 25-II headphones connected to the laptop. Each sound was a synthesised conga sound with a clear onset and rapid decay. Finger taps did not generate any auditory feedback (apart from a thud). Max/MSP controlled the presentation of all stimulus sequences and recorded tapping response times. Acoustic measurements taken prior to the experiment indicated that the mean end-to-end latency between a tap, registration of the tap by the MacBook and transmission of the sound through the headphones was approximately 28ms. This delay was accounted for in the Max/MSP script by subtracting 28ms from each registered tap onset time.

Three different types of auditory pacing signal were used. The regular pacing signal consisted of isochronous sequences of tones presented with inter-onset intervals (IOIs) of either 400ms (fast tempo) or 600ms (slow tempo). The tempo-changing signal contained tempo variations that mimic tempo changes in music as used by Pecenka and Keller (2011). The sequences progressed through both acceleration and deceleration of tempo following a sinusoidal function within a range of 400ms and 600ms. There were six such arrangements each one accelerating and decelerating at different time points throughout the tapping sequence.

Finally, the adaptive pacing signal or VP as used by Repp and Keller (2008) was used to obtain an estimate of participant's adaptive timing ability or 'human alpha' (HA). This adaptive sequence, responded to participant's tap timing by implementing phase correction to varying degrees. The parameter 'computer alpha' was set to four levels, namely 0, .3, .7, and 1. These figures represent the proportion of asynchrony that was corrected for in the subsequent tone in the sequence, specifically:

$t_{n+1} = t_n + T + a_c \times asy_n$,

$t_{n+1} = t_n + T + a_c \times asy_n$ $t_{n+1} = t_n + T + a_c \times asy_n$ where $t_n =$ = time of computer tone, $T =$ base IOI (either 400 or 600ms), $a_c =$ = phase correction parameter (0, .3, .7, or 1), and $asy_n =$ = asynchrony between tap and tone. For example, if a participant tapped too early (a negative asynchrony), the subsequent tone would occur earlier by a proportion (0, .3, .7, or 1) of the calculated negative asynchrony. Thus, each IOI throughout the sequence was adjusted in response to the amount and direction of participant tap asynchrony (see Figure 1). The adaptive sequences were presented at two base tempi: 400 and 600ms.

2.4. Procedure

Participants completed the test session individually within a quiet room in a laboratory. Participants were seated at a desk, upon which the MacBook and percussion pad were situated. Each individual was informed that they would hear several different types of tone sequences through the headphones and were asked to tap in time on the percussion pad with the sounds they heard. Instructions were presented via the MacBook. For the isochronous sequences the instructions were "using your finger, tap on the drum pad in time with the beat". For the tempo-changing participants were instructed to "keep in time with the changing tempo" whereas for the adaptive pacing sequences participants were asked to "keep in time with the beat while also maintaining a steady tempo, try to keep all of your taps even". Participants initially completed 11 practice trials of 20 tones each which included examples of all sequence types.

Participants then completed two test blocks of tapping trials, each comprised of 26 sequences of 60 tones each. All participants completed all trials in the same order, namely:

1. Four trials of the regular pacing sequence, alternating between 600 and 400ms.
2. The six tempo- changing sequences.
3. The 12 adaptive sequences inclusive of alpha = .3, .7, and 1, presented twice each at both 400 and 600ms.
4. A repeat of the four regular pacing sequences at both 400 and 600ms.

This test block took approximately 15 mins. After a short break, participants completed a second test block.

2.5. Data Analysis

Tapping data were pre-processed in a custom made MATLAB script. Linear interpolation was used to estimate the values of any missed taps. A missed tap may have been recorded due to a participant genuinely missing a tap or a tap not being registered by the drum pad. In addition, taps with a large asynchrony of > ± 200ms were also counted as missed taps. This criterion was chosen as 200ms represented 0.5 of the fastest IOI of 400ms. Any trials that were missing >5 out of 60 taps, or trials that included three or more consecutive missing taps were omitted from analysis. The criterion for statistical significance for all analyses was set to $a = .05$.

Synchronization Performance. Synchronization was assessed in terms of asynchronies, that is, the difference between tone onset times and the corresponding tap time. A negative asynchrony indicates that the tap preceded the tone. For each participant, three indicators of synchronization performance were calculated from the asynchronies in the isochronous trials. Synchronization *accuracy* was indexed by both mean signed asynchrony and mean absolute asynchrony. Signed asynchrony is an indicator of the relative earliness/ lateness of taps compared to the tones, whereas absolute asynchrony indicates the degree or magnitude of the asynchrony independent of the earliness or lateness. Synchronization *precision* was indexed by the coefficient of variation (CV; the SD of the signed asynchrony divided by the mean ITI for each trial). This is a measure of the variability of asynchrony, or how stable the tap timings were in relation to the tones. Higher numbers indicate more variability in tapping.

Prediction Indices. For each participant a prediction index (PI) was calculated using the data from the tempo changing trials. This index is a measure of the degree to which the participant was able to anticipate (predict) the tempo changes as opposed to following (tracking) the changes. An auto-regressive modelling approach was employed using R software. IOI and ITI time series data were first pre-whitened to remove autocorrelation (due to drift or local serial dependencies). Autoregressive coefficients representing the strength of the relationship between the participant's ITIs and the IOIs of the pacing tones were then calculated at both lag-0 and lag-1. The lag 0 coefficient indicates how accurately the participant anticipated or predicted the timing of the subsequent tone interval, whereas the lag1 coefficient reflects the degree to which the participant's tap intervals matched the timing of the previous tone interval (cf. Pecenka and Keller, 2011; Rankin et al., 2009; Repp, 2002). The lag 1 coefficient was then subtracted from the lag 0 coefficient, resulting in an index with values greater than 0 indicating higher degrees of predictive tendencies and values less than 0 indicating greater tracking tendencies. This procedure differs from that of Pecenka and Keller (2011), but the resulting estimates are strongly correlated with the prediction/tracking ratios computed using their method ($r = .98$).

We decided that a minimum of three valid tempo-changing sequences was required to arrive at reliable prediction indices. Sixteen participants did not meet this criterion and thus did not have a prediction index recorded and were excluded from further analysis.

Estimating human alpha. An estimate of each individual participant's use of phase correction, or human alpha, was calculated based on the lag-1 autocorrelations of the asynchronies (the time series correlated with itself at a lag of 1) across each level of alpha- 0, .3, .7, and 1 (alpha = 0 was calculated from the isochronous trials). A regression line was fitted to the mean autocorrelations for each level and the 0 crossing point (x- intercept) of this line of best fit was calculated. This represents the point of VP adaptivity that is optimal for each participant. This x- intercept was then subtracted from .9 (estimated optimal total error correction between the computer and human) to gain an estimate of human alpha (see Repp & Keller, 2008). Due to missing taps, we introduced the criterion that at least two valid trials for at least three levels of virtual partner alpha were required to calculate human alpha for an individual participant. One participant did not meet this criterion and thus did not have an alpha estimate recorded.

Table 1: Summary of descriptive statistics and intercorrelations between Human Alpha (HA) and Prediction Index (PI) with outcome measures of Mean Asynchrony, Mean Absolute Asynchrony and Coefficient of Variation.

Variables	Mean Asynchrony	M Absolute Asynchrony	Coefficient of Variation	Human Alpha	Prediction Index
HA	.52***	-.61***	-.76***	-	
PI	.34*	-.36**	-.65***	.54***	-
M	-56.65	56.78	.06	.31	.24
SD	25.02	24.50	.02	.22	.24

3. RESULTS

Analyses were conducted in SPSS. Prior to analysis, all data were screened for missing values and outliers. The 16 participants who did not have a prediction index calculated plus the 1 participant without a human alpha estimate recorded were excluded from analysis, leaving 52 participants in the final sample. No univariate outliers were identified (standardised scores in excess of ±3.29, Tabachnick & Fidell, 2007). Descriptive statistics can be seen in Table 1, while histograms displaying the distribution for each variable can be seen in Figure 2. The assumption of normality was violated for both Human Alpha Estimate (HA) and Coefficient of Variation (CV). This was rectified by logarithmic transformation of HA2 and square-root transformation of CV.

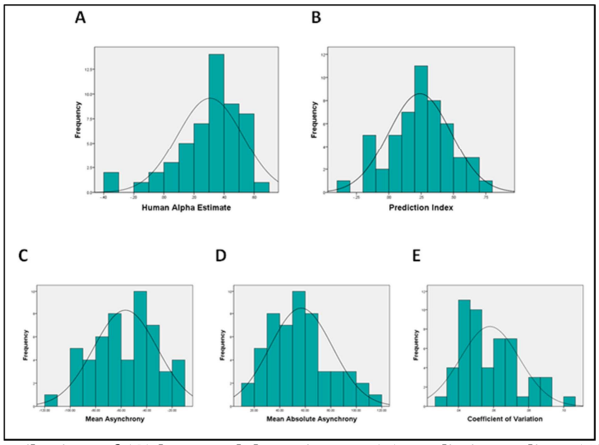


Figure 2: Histograms displaying the distributions of (A) human alpha estimates, (B) prediction Indices, (C) mean asynchrony, (D) mean absolute asynchrony, and (E) coefficient of variation.

3.1. Relationship between Adaptation and Anticipation

To investigate the relationship between anticipatory and adaptive mechanisms, a Pearson product-moment correlation was conducted between prediction indices (PI) and logarithmic transformed Human Alpha (logHA). There was a significant positive correlation $r(50) = .54$, $p = <.001$, indicating that higher error correction (human alpha) is related to higher predictive tendencies (see Figure 3).

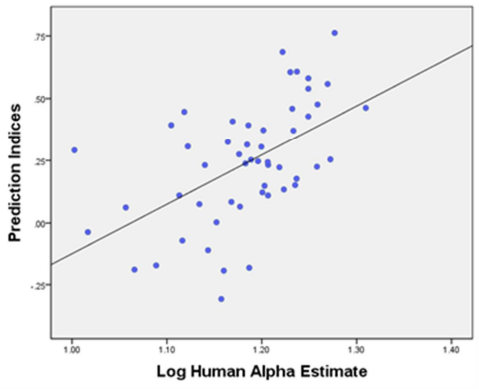


Figure 3: Scatterplot of the correlation between prediction indices and log human alpha estimate.

3.2. Predictors of Sensorimotor Synchronization Accuracy and Precision.

A series of hierarchical regressions were conducted in order to investigate the relationship between the combination of HA and PI, on each of the outcome variables of synchronization accuracy and coefficient of variation [sqrtCV]. The assumptions of multiple regression were evaluated, along with assessment of multivariate outliers and multicollinearity between predictors. No multivariate outliers were identified (no standardised residual >3 or extreme Mahalanobis distance score $p < .001$, critical value = 13.82).

In all three regressions, logHA was entered first, followed by PI. This order represents the order of interest in each variable, with HA being first due to this being a more automatic process. Table 1 shows intercorrelations among the predictor variables and each dependant

2 Due to the negative skewness of HA, this variable was reflected before transformation and then the reflection was reversed (Tabachnick & Fidell, 2007)

measure of asynchrony, absolute asynchrony, and sqrtCV. While Table 2 displays the standardised regression coefficients (β) and R^2 change (ΔR^2) for the predictors within each regression.

Mean asynchrony—one measure of synchronization accuracy—was the outcome variable of the first regression. Evaluation of the overall multiple regression revealed that at step 1, HA alone was a significant predictor R = .52, R^2 = .27, adjusted R^2 = .26, F (1, 50) = 18.76, p = >.001. The addition of PI at step 2 did not significantly enhance the predictive utility of the final model R = .53, R^2 = .28, adjusted R^2 = .25, F (2, 49) = 9.42, p = >.001, nor was PI a significant unique predictor of mean asynchrony once shared variance was removed (see Table 2). A second hierarchical regression was performed with mean absolute asynchrony—a second measure of accuracy—as the dependant measure. This regression mirrors the previous model with HA entered at step 1 being sufficient to generate a significant model R = .61, R^2 = .38, adjusted R^2 = .36, F (1, 50) = 29.98, p < .001. PI entered at step 2 is again not a significant unique predictor (see Table 2) and does not improve the final model R = .61, R^2 = .38, adjusted R^2 = .35, F (2, 49) = 14.18, p < .001.

Table 2: Hierarchical Multiple Regression of Human Alpha and Prediction Index as Predictors of Asynchrony, Absolute Asynchrony and Coefficient of Variation.

	Mean Asynchrony		Mean Absolute Asynchrony		Coefficient of Variation	
Variables	β	ΔR^2	β	ΔR^2	β	ΔR^2
Step 1		.27***		.38***		.57***
Log HA	.52***		-.61***		-.76***	
Step 2		.01		.01		.09**
Log HA	.48**		-.59***		-.56***	
PI	.08		-.05		-.36**	

Finally, to assess the best predictors of tapping precision, a third hierarchical regression was conducted with sqrtCV as the dependant measure. Again at step 1, HA alone generated a significant regression model R = .76, R^2 = .57, adjusted R^2 = .57, F (1, 50) = 66.34, p < .001. Unlike the previous regressions, PI significantly improved the model when entered at step 2, R = .81, R^2 = .66, adjusted R^2 = .65, F (2, 49) = 47.92, p < .001, and is a significant unique predictor of tapping variability. This model accounts for approximately 66% of sample variance in tapping precision.

As expected, HA was a significant predictor in all the three regressions, and was sufficient to explain a significant amount of variance in all three dependent tapping measures. Interestingly, once shared variance was removed, PI was only a significant unique predictor of sqrtCV, the measure of tapping precision, and did not significantly contribute to either of the dependant measures of synchronization accuracy. This is despite PI having a significant bivariate correlation with both mean asynchrony and absolute asynchrony. Out of the three regressions, the most variance explained was with tapping precision (coefficient of variation), followed by absolute asynchrony, and the least was with signed asynchrony. Together these results indicate that while adaptive mechanisms are important for both synchronization accuracy and precision, anticipatory mechanisms are more strongly related to synchronization precision than synchronization accuracy.

4. DISCUSSION

The relationship between two of the underlying mechanisms of synchronization, anticipation and adaptive timing, was investigated by assessing the correlation between human alpha (HA) estimates and prediction index (PI) scores. A positive correlation was found between these two measures, indicating that higher degrees of phase correction are related to higher tendencies to predict timing of upcoming sounds. These results suggest that these two processes are not independent but instead work in tandem to improve synchronization performance. The positive relationship implies that, without disrupting timing stability, higher prediction tendencies allow more error correction due to correcting for a larger proportion of smaller asynchronies. In other words, higher prediction leads to higher accuracy (i.e. smaller asynchronies; Pecenka & Keller, 2011), so the absolute amount of correction is small even when a greater relative proportion of the asynchrony is corrected for.

The observed positive relationship between anticipatory and adaptive mechanisms is a first step towards understanding the interrelationship between the underlying mechanisms of sensorimotor synchronization. Though no causal inferences can be made from this study (one may be influencing the other, or a separate mechanism may be driving both), it is hypothesised that the link is the internal models that run online simulations of actions before they occur. This process within the sensorimotor control system appears to access both mechanisms of anticipating the timing of the subsequent sound, and monitoring and correcting synchronization errors. It is has been claimed that this form of anticipatory error correction is implemented by 'joint' internals models that compute timing errors prospectively based on simulations of the predicted timing of another's action and the timing of one's own next action given a certain error correction setting (van der Steen & Keller, 2013).

The secondary aim was to assess the contribution of anticipatory and adaptive mechanisms to individual differences in synchronization precision and accuracy. As expected, evaluation of the hierarchical multiple regression analyses revealed that together the combination of phase correction and anticipation accounted for a significant amount of the variance in each of the dependant measures of accuracy (mean signed asynchrony and mean absolute asynchrony) and precision (coefficient of variation). This combination

of variables best accounted for the variance in the precision measure (CV), accounting for 66% variance in this sample. Despite each of the overall models being significant, phase correction was the only variable to contribute significantly in all three regressions. Contrary to expectations, the addition of prediction indices did not significantly improve either of the models that assessed synchronization accuracy. In addition, once shared variance was removed, the prediction index was only a significant unique predictor of tapping precision. This indicates that anticipatory tendencies may be uniquely related more so to the stability of inter-tap intervals rather than reducing asynchrony.

The findings of the present study confirm that individual differences in both anticipatory and adaptive mechanisms influence synchronization performance. The present results buttress those of Pecenka and Keller (2011), who found that anticipatory tendencies are predictive of both accuracy and precision. The current findings confirm and extend these results by adding assessment of adaptive timing. We demonstrated that anticipatory tendencies are related to precision and accuracy, and that anticipation explains variance above and beyond that explained by adaptive timing in the precision measure. However, we also found that in the case of synchronization accuracy, the variance explained by anticipation is shared with adaptive timing, and does not explain anything further than adaptive timing alone.

Similar to Fairhurst et al. (2012) and Repp and Keller (2008), the results of this study also further validate the use of the virtual partner as a valid tool for investigating dynamic sensorimotor synchronization. Manipulation of the parameter 'alpha' is a simple but effective means to simulate varying synchronization ability. This paradigm can be used as a tool to assist people with irregular or disrupted rhythmic timing. For example, Hove, Suzuki, Uchitomi, Orimo, and Miyake (2012) have used a related form of adaptive pacing signal to assist Parkinson's disease patients to correct gait irregularities and instability. The VP may also be used to study populations that have difficulty with social coordination tasks, such as those with deficits in social processes (e.g. autism spectrum disorders). Likewise, this tool may be used to investigate early developmental processes and the impact of aging and cognitive decline on sensorimotor synchronization.

In conclusion, this study has identified a relationship between anticipatory and adaptive processes, two of the underlying mechanisms of sensorimotor synchronization. In addition, the results demonstrate that individual differences in these two mechanisms can predict both synchronization accuracy and precision. The present study has also demonstrated further support for the VP as a valid tool for identifying error correction tendencies and investigating the dynamics of sensorimotor synchronization. A promising avenue in ongoing research is to use computational modelling approaches (van der Steen & Keller, 2013) to further explain the nature of the relationship between anticipation and adaptation. In addition, research into sensorimotor synchronization ability within special populations may lead to identification of further ways that this paradigm may benefit those with disrupted rhythmic timing abilities.

5. ACKNOWLEDGEMENTS

The Authors would like to thank Dr Ben Shultz for assistance with MATLAB programming and pre-processing the data.

6. REFERENCES

Aschersleben, G., Stenneken, P., Cole, J., & Prinz, W. (2002). Timing mechanisms in sensorimotor synchronisation. In Prinz, W., & Hommel, B. (Eds.), Common mechanisms in perception and action: Attention and performance XIX (pp. 227- 244). Oxford: Oxford University Press.

Fairhurst, M. T., Janata, P., & Keller, P. E. (2012). Being and feeling in sync with an adaptive virtual partner: brain mechanisms underlying dynamic cooperativity. Cerebral Cortex.

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, F., A. Carassa, & G. Riva (Eds.), Enacting intersubjectivity: A cognitive and social perspective on the study of interactions (pp. 205-221). Amsterdam: IOS Press.

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), Expressiveness in music performance: Empirical approaches across styles and cultures (pp.260-282). Oxford: Oxford University Press.

Keller, P. E. (2012). Mental imagery in music performance: underlying mechanisms and potential benefits. Annals of the New York Academy of Sciences, 1252. 206-213. doi: 10.1111/j.1749-6632.2011.06439.x

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), Expressiveness in music performance: Empirical approaches across styles and cultures (pp. 260-282). Oxford: Oxford University Press.

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. Music Perception: An Interdisciplinary Journal, 28, 27-46. doi: 10.1525/mp.2010.28.1.27

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. Consciousness and Cognition, 16, 102-111. doi: 10.1016/j.concog.2005.12.004

Knoblich, G., Butterfill, D., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. In: Ross B (Ed.), The psychology of learning and motivation, vol 54. (pp. 59 – 101). Burlington: Academic Press.

Goebl, W. & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. Music Perception, 26, 5, 427 – 438. doi: 10.1525/M.P.2009.26.5.427

Hove, M. J., Suzuki, K., Uchitomi, H., Orimo, S., & Miyake, Y. (2012). Interactive Rhythmic Auditory Stimulation Reinstates Natural 1/f Timing in Gait of Parkinson's Patients. PlosOne, 7, 1-8. doi:10.1371/journal.pone.0032600

Mates, J. (1994). A model of synchronization of motor acts to a stimulus sequence. Biological cybernetics, 70(5), 463-473.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. Cognition, 107(1), 179-217.

Pecenka, N., Engel, A., & Keller, P.E. (2013). Neural correlates of auditory temporal predictions during sensorimotor synchronization. Frontiers in Human Neuroscience, 7: 380.

Pecenka, N., & Keller, P. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. Experimental Brain Research, 211, 505 – 515. doi. 10.1007/s00221-011-2616-0

Phillips-Silver, J., & Keller, P.E. (2012). Searching for roots of entrainment and joint action in early musical interactions. Frontiers in Human Neuroscience, 6:26. doi: 10.3389/fnhum.2012.00026

Rankin, S. K., Large, E. W., & Fink, P. W. (2009). Fractal tempo fluctuation and pulse prediction. Music Perception, 26, 401-413. doi: 10.1525/MP.2009.26.5.401

Repp, B. H. (2001). Processes underlying adaptation to tempo changes in sensorimotor synchronization. Human movement science, 20(3), 277-312.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. Psychonomic Bulletin and Review, 12, 6, 969 – 992. doi: 10.3758/BF03206433.

Repp, B. H. (2011). Tapping in synchrony with a perturbed metronome: The phase correction response to small and large phase shifts as a function of tempo. Journal of Motor Behavior, 43, 213-227. doi: 10.1080/00222895.2011.561377

Repp, B. H., & Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. Quarterly Journal of Experimental Psychology Section A, 57(3), 499-521.

Repp, B. H. & Keller, P. (2008). Sensorimotor synchronization with adaptively timed sequences. Human Movement Sciences, 27, 423 – 456. doi: 10.1016/j.humov.2008.02.016.

Repp, B. H. & Su, Y. H. (2013). Sensorimotor Synchronisation: A review of recent research (2006 – 2012). Psychonomic Bulletin and Review, 20, 403-452. doi: 10.3758/s13423-012-0371-2.

Stoit, A., van Schie, H. T., Riem, M., Meulenbroek, R. G., Newman-Norlund, R. D., Slaats-Willemse, D. I., ... & Buitelaar, J. K. (2011). Internal model deficits impair joint action in children and adolescents with autism spectrum disorders. Research in Autism Spectrum Disorders, 5(4), 1526-1537.

Van der Steen, M. C., & Keller, P.E. (2013). The adaptation and anticipation model (ADAM) of sensorimotor synchronization. Fronteirs in Human Neuroscience. 7: 253. doi: 10.3389%2Ffnhum.2013.00253

Vorberg, D., & Schulze, H. H. (2002). Linear phase-correction in synchronization: Predictions, parameter estimation, and simulations. Journal of Mathematical Psychology, 46(1), 56-87.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. Trends in cognitive sciences, 2(9), 338-347.

EFFECTS OF MUSICAL EXPERTISE ON AUDIOVISUAL INTEGRATION: INSTRUMENT-SPECIFIC OR GENERALISABLE?

Laura Bishop[1], Werner Goebl[2]
[1]Austrian Research Institute for Artificial Intelligence, Austria, [2]Institute of Music Acoustics, University of Music and Performing Arts Vienna, Austria
Time: 18:00

1. Background: During ensemble performance, musicians exchange auditory and visual signals that can help in synchronising with each other's actions. Musicians must integrate corresponding auditory and visual signals accurately to make use of them. Precision in audiovisual integration improves with increasing perceptual-motor expertise, perhaps because experts are better able to predict when the auditory effects of observed actions should occur. Performance expertise has been found to have a greater effect than visual expertise on musicians' prediction abilities during synchronisation tasks, with performers better able to synchronise with actions that fall within their own motor repertoires than with actions they have only ever observed. It is unclear whether the effects of expertise on audiovisual integration are likewise instrument-specific or generalisable across instruments, however.

2. Aims: The present study investigated the potential instrument-specific effects of expertise on audiovisual integration. Expertise in playing a particular instrument was hypothesised to facilitate prediction of observed actions, increasing sensitivity to audiovisual asynchrony in that instrumental context.

3. Method: Ten-second clips were extracted from audio-video recordings of clarinet, piano, and violin performances, and presented to highly-skilled clarinettists, pianists, and violinists. Clips either maintained the audiovisual synchrony present in the original recording or were modified so that the video led or lagged behind the audio. Participants indicated as quickly as possible whether the audio and video channels in each clip were synchronised. Sensitivity to audiovisual asynchrony was assessed for each expertise group/stimulus instrument pairing by evaluating the mean point of subjective synchrony and the mean range of asynchronies most often rated as synchronised (i.e. temporal integration window; TIW).

4. Results: Though participants across expertise groups detected asynchronies most readily in piano and least readily in violin stimuli, pianists performed significantly better for piano than for clarinet or violin. A relationship between musical training and TIW was also observed with data pooled across stimuli. Thus, sensitivity to audiovisual asynchrony improved generally with increasing expertise, and only pianists showed facilitation for their own instrument.

5. Conclusions: Sensitivity to audiovisual asynchrony was affected by musical training and the nature of sound-producing movements observed. The results suggest that, to some extent, the effects of performance expertise can be instrument-specific, though they may generalise across instrumental contexts more readily during audiovisual asynchrony detection tasks than during synchronisation tasks, when overt, precisely-timed movements are required.

[2E] When Music Delivers Pleasure

LOS ANGELES 16:00-18:30 Monday 4 Aug 2014

"DELAY" IN CLIMAX: CONTEXTUALIZATION, ANALYSIS, AND IMPLICATION

Ji Yeon Lee[1]
[1]*City University of New York, Graduate Center, USA*

Time: 16:30

When a winner of the grand prix is announced at competition or award ceremony, an announcement of the final outcome is treated to a tantalizing delay, the moderator's pauses often accompanied by a drum roll. This produces maximal expectation in the audience and dramatizes the revelation as a long-awaited culmination. In "Sweet Anticipation" (2008), David Huron notes the emotional effect of such delay; the present paper departs from Huron's idea of delay and attempts to deepen and calibrate the topic within a music theoretical perspective through contextualization, analysis, and implication. I position delay in the context of a climax system comprised of four stages: initiation via tension onset, intensification through pace acceleration and tension enhancement, highpoint as arrival of an expected goal, and abatement through discharge of the accumulated tension. The stage of intensification sometimes concludes with delay maximizing expectation; delay increases listeners' emotional suspension by postponing highpoint, parallel to a final hesitation or breath before arrival at a goal. Based on the climax system, I will probe delay examples drawn from a wide range of the Romantic repertoire: etudes by Liszt and Rachmaninoff, and operas by Wagner and verismo composers. I suggest three parameters for delay formation as analytical methodology: (1) score indications--ritardando, rallentando, allargando--that dictate pace for the performers; (2) musico-rhetorical devices including pause, fermata, sudden change to soft dynamic, and extended repetition of notes or chords; (3) devices embodied within the musical structure--temporal increase of the phrase group unit, prolongation of a chord via neighbor chords and applied dominants, and deceptive cadences. This study will treat the subjective, intuitive feeling of delay to a systematic analysis of the mechanisms used to create it in musical practice. Furthermore, the study will contribute to performance practice by offering specific examples of delay, illuminating how musico-rhetorical and structural devices suggest tempo adjustment to performers without score indications.

WHEN THE MUSICAL REWARD ARRIVES: LISTENERS' PHYSIOLOGICAL RESPONSES TO THE THEME ENTRANCE/RECURRENCE IN THE VERSE-CHORUS FORM AND THE SONATA FORM

Chen-Gia Tsai[1], *Chung-Ping Chen*[1], *Szu-Pei Yu*[2]
[1]*Graduate Institute of Musicology, National Taiwan University, Taiwan,* [2]*Department of Biochemical Science and Technology, National Taiwan University, Taiwan*

Time: 17:00

Pleasurable music serves as an abstract reward to listeners, and it is recently demonstrated that the 'wanting' phase and the 'liking' phase were associated with distinct activity in the brain's reward system [Salimpoor et al. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. Nature Neuroscience 14:257-262]. The present study examines the rewarding nature of musical themes in the verse-chorus form and the classical sonata form using two physiological measures: skin conductance response (SCR) and finger temperature (FT). Stimuli of the experiment of popular songs were the verse-chorus progressions of 15 popular songs. We found significant ($p<.0001$, uncorrected) SCRs at the entrance of the chorus. These responses indicate that the verse-to-chorus transition transiently elevated listeners' arousal, possibly reflecting listeners' intense anticipation immediately before the chorus. In the experiment of sonata form, participants listened to the development-recapitulation progression of six musical works by Beethoven and Mozart. Listeners' SCRs induced by the theme recurrence tend to occur at the end of its theme's first phrase. Moreover, we found FT a good indicator of the relaxation level and emotional valence. For Mozart's Symphony No. 40 and Beethoven's Symphony No. 7, FT significantly ($p<.05$) increased at the goal-directing concluding passage of the development (also known as the retransition). In contrast, the 'tragic' retransition in the first movement in Beethoven's Symphony No. 5 significantly ($p<.05$) decreased listeners' FT. The first movement of Beethoven's Piano Concerto No. 5 provides an interesting example of the preparation of theme recurrence. The first part of its retransition sustains a diminished seventh chord with a slowly climbing, vanishing piano arpeggios. This gesture resulted in an increase in FT and an anti-peak of SCR amplitude, followed by two SCR peaks at the preparation and the arrival of the major theme. In sum, we found that SCRs reflect listeners' elevated arousal associated with (1) the feeling of wanting for anticipating the chorus in popular songs, and (2) the feeling of liking for recognizing the major theme at the beginning of the recapitulation in sonata form. On the other hand, listeners' emotion valence may be reflected by FT.

MUSIC WITH VARYING DEGREES OF COMPLEXITY ARE RATED DIFFERENTLY ON LIKING AND COMPLEXITY BY MEN COMPARED TO WOMEN.

Johan Paulin[1], *Guy Madison*[1]
[1]*Department of Psychology, Umeå University, Sweden*

Time: 17:30

ABSTRACT

Music is a human universal, practiced in all cultures and continents throughout recorded history. It is clear that performing and listening to music taps into numerous areas of our brain and to excel as it requires large amounts of practice, thus it is a skill that's hard to

fake and from a survival point of view shows few advantages. One possible evolutionary explanation for the longevity of music is that it might have evolved as a way to display genetic fitness and mating value. Among other things, this would manifest as sex differences for complex music when it comes to appreciation.

In this study we tested if there was a sex difference when it comes to Liking as well as how accurate participants were at being rating complexity in music. Results and potential avenues for further research are discussed

1. BACKGROUND

Music is a human universal, practiced in all cultures and continents throughout recorded history (Pinker, 2002). Despite this, music does not seem to have any apparent survival value. On the contrary, music-making and practice, while easy in the beginning, often has a steep learning curve and requires much dedication and resources to become skilled at. Resources that at first glance might be better spent on some more dire activity. Music is also notoriously loud, making it a less ideal activity if wanting to avoid predators or hostile strangers. None the less, given its commonality among humans, it seems plausible that music has some kind of evolutionary benefit to it, however obscure, that has been passed down through the millennia.

One such possible benefit might be if music-making evolved as a "costly signal" of the individual's genetic fitness and high mating value (Miller, 2000). Costly signaling-theory states that qualities that are not easily detectable but are linked to an organism's fitness, are communicated through signals that imply a high cost to produce for the organism (Zahavi, 1975). A peacock's tail is for example cumbersome and requires a lot of resources to grow, thus it is only the fittest, healthiest males who develop the most extravagant tails, thereby avoiding fakes. Music-making has been shown to positively correlate with intelligence (Ruthsatz, Detterman, Griscom, & Cirullo, 2008), higher cognitive functions (Zatorre, 2005), improved motor timing (Baer, Thibodeau, Gralnick, Li, & Penhune, 2013) and well-being (Klassen, Liang, Tjosvold, Klassen, & Hartling, 2008; MacDonald, Kreutz, & Mitchell, 2012), there for it is reasonable to consider that music might be an underlying signal of high genetic fit. Given this, men should show a preference for music with a high structural complexity sine it would be a suitable display of fitness due to the amount of resources needed to excel at it, in conjuncture with it being difficult to fake.

In this study we investigated if there is a sex difference when it comes to liking music with varying degrees of complexity. We hypothesize that men and women rate ecologically valid music differently with respect to Complexity and Liking. More specifically; men will show a higher liking of complex music, and more accurately gauge complexity in music compared to women.

2. METHOD

132 participants (78 women) were randomly recruited in the vicinity of a medium sized Swedish city, using flyers, internet-promotion and word of mouth. Background variables such as music preference were controlled for.

The music excerpts (ME) were lifted from commercially available recordings of varying popular genres such as pop, rock, metal, electronica and folk music, and edited down to 15 second clips. ME:s were then screened for complexity by a group of music experts with many years of listening experience (Years of music listening: M = 30,57, SD = 11,25; Hours per week of music listening: M = 22, SD = 11,6). The expert raters showed within-correlations of .80 on their ratings of the ME, indicating high consistency of inter-rater agreement. From their ratings 55 ME:s were selected on five levels of complexity with 11 ME:s in each level.

To avoid order effects, all participants listened to the stimuli in a randomized order. Ratings of Complexity and Liking of the ME:s were obtained using a 11-point Likert-scale.

3. RESULTS

Multivariate analysis of variance (MANOVA) were performed using Sex as independent variable and Liking and Complexity as dependent variables. There were a significant effect of Sex, Wilks' $\lambda = .993$, F (2, 7366) = 27,53, $p < .000$.

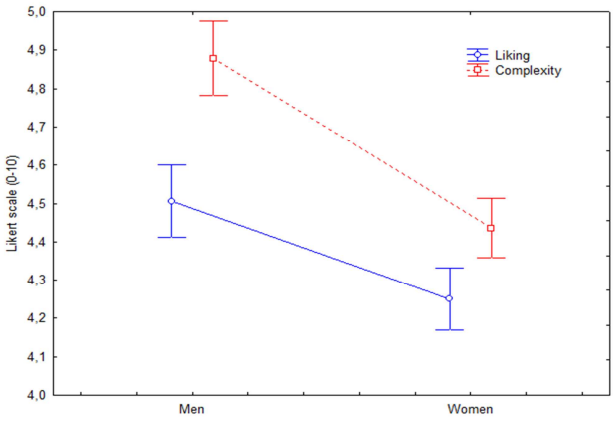


Figure 1: Mean ratings and conf. intervals (95%) of Liking and Complexity as a function of Sex

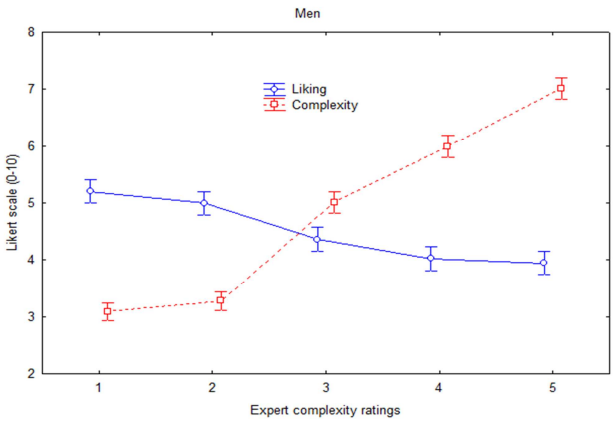


Figure 2: Mean ratings and conf. intervals (95%) of Liking and Complexity for men as a function of Expert complexity ratings.

Given the significance of the overall test we performed separate Univariate analysis of variance (ANOVA) for the DV:s. There was significant main effects of both Liking: F (1, 7367) = 15,790, p < .000 and Complexity: F (1, 7367) = 49,756, p < .000. Figure 1 (below) illustrates the mean ratings of the DV:s as a function of Sex.

To see the ratings of men (fig. 2) and women (fig. 3) separately, plots for Liking and Complexity were made using the experts' complexity ratings as IV.

4. DISCUSSION

Based on the hypothesis that music can act as a costly signal of genetic fitness we tested if men and women differed in a) discriminating between, and b) their appreciation of, complex music. We found a significant effect of sex on both Liking and Complexity, indicating that men and women did indeed rate the music examples differently.

Men used a broader range when rating Complexity, compared to women, indicating a greater ability to discriminate between more and less complex music. For liking, both sexes showed a negative trend the more complex the ME:s were, however while men showed very little change in Liking between the least and the most complex music categories, women showed a more radical shift in dislike the more complex the ME:s became. This seems to suggest that men's appreciation were less disturbed by the increasing complexity of the music than women.

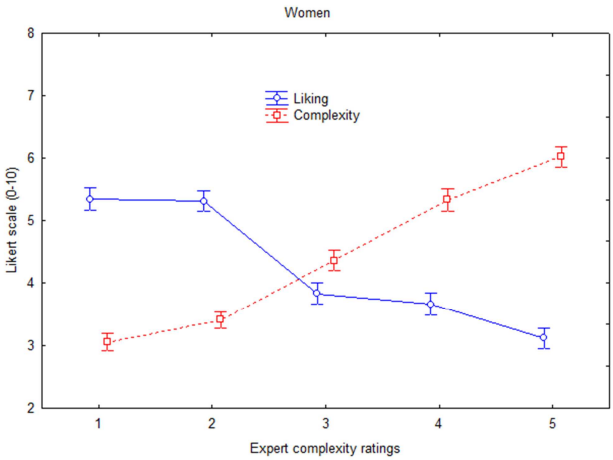


Figure 3: Mean ratings and conf. intervals (95%) of Liking and Complexity for women as a function of Expert complexity ratings.

While these results do not prove or disprove Millers (2000) theory of music as a result of sexual selection, it does show that there is a difference, which in turn might point towards deeper, more hardwired differences due to selection pressures. For future research it would be desirable to see if this sex difference is limited to biological sex or if for example people exposed to high levels of prenatal androgens, regard less of sex, perceive complex music differently than people with low exposure. This prenatal exposure, as measured by the ratio of the length of the second and fourth digit (Galis, Ten Broek, Van Dongen, & Wijnaendts, 2010) has been linked to other factors such as cognitive ability (John T. Manning, Baron-Cohen, Wheelwright, & Fink, 2010), verbal fluency, personality traits (Neave, Laing, Fink, & Manning, 2003), and physical ability (Hönekopp, T Manning, & Müller, 2006; J T Manning, Scutt, Wilson, & Lewis-Jones, 1998;; John T. Manning & Fink, 2011). A recent study (Crocchiola, 2014) involving visual artists showed that successful male and female artist had lower 2D:4D (higher exposure to prenatal androgen) than controls; supporting Millers theory.

5. REFERENCES

Baer, L. H., Thibodeau, J. L. N., Gralnick, T. M., Li, K. Z. H., & Penhune, V. B. (2013). The role of musical training in emergent and event-based timing. Frontiers in Human Neuroscience, 7*(May), 191. doi:10.3389/fnhum.2013.00191*

Crocchiola, D. (2014). Art as an Indicator of Male Fitness : Does Prenatal Testosterone Influence Artistic Ability ? Evolutionary Psychology, 12*(3), 521–533.*

Galis, F., Ten Broek, C. M. a, Van Dongen, S., & Wijnaendts, L. C. D. (2010). Sexual dimorphism in the prenatal digit ratio (2D:4D). Archives of Sexual Behavior, 39*(1), 57–62. doi:10.1007/s10508-009-9485-7*

Hönekopp, J., T Manning, J., & Müller, C. (2006). Digit ratio (2D:4D) and physical fitness in males and females: Evidence for effects of prenatal androgens on sexually selected traits. Hormones and Behavior, 49*(4), 545–9. doi:10.1016/j.yhbeh.2005.11.006*

Klassen, J. A., Liang, Y., Tjosvold, L., Klassen, T. P., & Hartling, L. (2008). Music for pain and anxiety in children undergoing medical procedures: a systematic review of randomized controlled trials. Ambulatory Pediatrics : The Official Journal of the Ambulatory Pediatric Association, 8*(2), 117–28. doi:10.1016/j.ambp.2007.12.005*

MacDonald, R., Kreutz, G., & Mitchell, L. (2012). Music, health, and wellbeing*. Oxford University Press.*

Manning, J. T., Baron-Cohen, S., Wheelwright, S., & Fink, B. (2010). Is digit ratio (2D:4D) related to systemizing and empathizing? Evidence from direct finger measurements reported in the BBC internet survey. Personality and Individual Differences, 48*(6), 767–771. doi:10.1016/j.paid.2010.01.030*

Manning, J. T., & Fink, B. (2011). Digit ratio (2D:4D) and aggregate personality scores across nations: Data from the BBC internet study. Personality and Individual Differences, 51*(4), 387–391. doi:10.1016/j.paid.2010.05.040*

Manning, J. T., Scutt, D., Wilson, J., & Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. Human Reproduction , 13 *(11), 3000–3004. doi:10.1093/humrep/13.11.3000*

Miller, G. F. (2000). The Mating Mind: How Sexual Choice Shaped The Evolution Of Human Nature *(p. 503). New York: Anchor books.*

Neave, N., Laing, S., Fink, B., & Manning, J. T. (2003). Second to fourth digit ratio, testosterone and perceived male dominance. Proceedings. Biological Sciences / The Royal Society, 270*(1529), 2167–72. doi:10.1098/rspb.2003.2502*

Pinker, S. (2002). The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.*

Ruthsatz, J., Detterman, D., Griscom, W. S., & Cirullo, B. a. (2008). Becoming an expert in the musical domain: It takes more than just practice. Intelligence, 36*(4), 330–338. doi:10.1016/j.intell.2007.08.003*

Zahavi, A. (1975). Mate selection-a selection for a handicap. Journal of Theoretical Biology, 53*(1), 205–14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1195756*

Zatorre, R. (2005). Music, the food of neuroscience? Nature, 434(March), 312–315. Retrieved from http://www.nature.com/nature/journal/v434/n7031/full/434312a.html

Session 3

[3A] Cognitive Neuroscience of Music 1

SEOUL 08:30- 10:30 Tuesday 5 Aug 2014

ELECTROPHYSIOLOGICAL CORRELATES OF MUSICAL TEMPO PREFERENCES

Anna-Katharina Raphaela Bauer[1], Gunter Kreutz[2], Christoph Herrmann[3,4]
[1]Neuropsychology Lab, Department of Psychology, Carl Von Ossietzky University, Oldenburg, Germany, [2]Department of Music, Carl Von Ossietzky University, Oldenburg, Germany, [3]Experimental Psychology Lab, Department of Psychology, Cluster of Excellence "hearing4all", European Medical School, Carl Von Ossietzky University, Oldenburg, Germany, [4]Research Center Neurosensory Science, Carl Von Ossietzky University, Oldenburg, Germany

Time: 08:30

Preferred musical tempo peaks slightly above 120 beats per minute and is subject to inter-individual variation. The preferred tempo is believed to be associated with rhythmic body movements as well as motor cortex activity. However, a long-standing question is whether preferred tempo is determined biologically. We first aimed to examine an individual's preferred tempo using a multi-step procedure. Subsequently, we correlated the preferred tempo to perceptual as well as motor EEG correlates – namely individual alpha frequency (IAF), evoked gamma band response (eGBR), and motor beta activity (MB). A subsidiary goal was to assess the mediating effects of musicality (Gold-MSI) on preferred tempo. Over three consecutive days the preferred tempo of 12 normal-hearing subjects aged 21–28 years (6 female, M=25.0, SD=2.32) was determined using a two-step approach, combining both perceptual and motor measurements. While listening to rock music excerpts, we first examined an estimate of an individual's preferred tempo using a mere perceptual task, followed by a more accurate examination of preferred tempo by combining perceptual and motor aspects. Further, in three music unrelated tasks the IAF, eGBR, and MB activity was recorded using a BrainAmp system with 32 Ag/AgCl sintered electrodes, placed according to the 10-10 system. Further, mood (PANAS) and genre preference (STOMP) were assessed as covariates. Musicality ranged from non-musicians to well-trained musicians (Gold-MSI General Musical Sophistication score: range 44 – 110, M = 73, SD = 23). One participant had to be excluded from analysis due to a lack of tempo perception ability, as indexed by a simple linear regression. The correlation between the estimated and accurate preferred tempo was highly significant (Pearson's r(9) = .79, p = .004). Further, the EEG data revealed a positive correlation between preferred tempo and MB activity (Pearson's r(9) = .71, p = .014). No significant correlations between preferred tempo and IAF or eGBR were found. Musicality had no influence on preferred tempo, but one facet of musicality, perceptual abilities, could explain about 29% of the variance in IAF. The results demonstrate a link between preferred tempo and neural motor activity but no relation to perceptual EEG correlates (IAF, eGBR) was found. These findings suggest that individual tempo preferences result from neural activity in the motor cortex which shows inter-individual variation. More research is however needed to generalize the present findings across different styles of music.

THE ROLE OF THE INFERIOR FRONTAL GYRUS IN BEAT AND METRE PERCEPTION

Sujin Hong[1], Neil Roberts[2], Alexa Morcom[3], Katie Overy[1]
[1]Institute for Music and Human Social Development (IMHSD), Reid School of Music, Edinburgh College of Art, The University of Edinburgh, United Kingdom, [2]Clinical Research Imaging Ventre (CRIC), The University of Edinburgh, United Kingdom, [3]Department of Psychology, The University of Edinburgh, United Kingdom

Time: 09:00

Previous studies into the neural correlates of temporal regularity in musical rhythm have consistently shown activation in auditory, premotor (PM), supplementary motor areas (SMA), and cerebellum. However, the functional role of frontal regions in particular the inferior frontal gyrus (IFG), a classic language and motor sequencing region, is currently less well defined, although co-activation between the IFG and the auditory and sensorimotor areas has been shown for rhythm processing tasks. We investigated the role of IFG during beat and metre perception, using functional Magnetic Resonance Imaging (fMRI). In Study 1, the temporal regularity of a series of rhythm pairs was either enhanced or weakened by the addition of volume accents (e.g. Unaccented, Beat Accented or Non-Beat Accented rhythm pairs). Participants (N14, 6 males) listened to two successive rhythms as a pair, and judged whether the rhythms were the same or different. The results showed that, compared with Beat Accented rhythms, Non-Beat Accented rhythms significantly activated left IFG while all conditions activated right IFG. In Study 2, a series of isochronous beats were metrically organised into groups of 2/4, 3/4 or 4/4 using pitch accents. Participants (N15, 6 males) listened to the steady beat stimuli, and judged the type of metrical organisation. The results showed that activation in bilateral IFG increased parametrically in accordance with metrical organisation levels (i.e. from 2/4 to 3/4 to 4/4). In both studies, we also found extensive activation across auditory and sensorimotor areas including the primary and secondary auditory areas, PM, SMA, and basal ganglia, consistent with previous studies. The results of these two experiments support previous findings regarding the importance of auditory-motor networks, and provide converging evidence for a functional role of bilateral IFG in two key aspects of musical organisation, with the IFG not only involved in perceiving temporal regularity, but also in the processing of higher level groupings of metre.

DURATION PERCEPTION OF FILLED AND EMPTY INTERVALS—A STUDY WITH MAGNITUDE ESTIMATION AND ELECTROENCEPHALOGRAPHY

Emi Hasuo[1,2], Kazuo Ueda[3], Takuya Kishida[3], Haruna Fujihira[3], Satoshi Morimoto[3], Gerard B. Remijn[3], Kimio Shiraishi[3], Shozo Tobimatsu[1], Yoshitaka Nakajima[3]

[1]Faculty of Medical Sciences, Kyushu University, Japan, [2]Japan Society for the Promotion of Science, Japan, [3]Faculty/Graduate School of Design, Kyushu University, Japan
Time: 09:30

Durations of sounds and silences between sounds play important roles in music rhythm. The duration between the onset and the offset of a sustained sound is called a filled interval, and the duration between the onsets of two very short sounds is called an empty interval. A filled interval is often perceived to be longer than an empty interval of the same physical duration in psychophysical experiments, but the occurrence of this overestimation seemed to depend on the duration range. Our aim was to examine the perceived durations of empty and filled intervals to explore whether their relationship differed for shorter and for longer intervals. We measured the subjective durations of filled and empty time intervals of 30, 49, 81, 134, 221, 364, and 600 ms by using magnitude estimation. Time intervals were presented in random order, and participants responded to each presentation with a positive number corresponding to the perceived duration of the presented time interval. The geometric mean of the responses from 41 participants was calculated for each stimulus condition. We recorded event related potentials (ERPs) in 15 of these participants while they listened to filled and empty intervals of 221 and 600 ms. The results for the magnitude estimation experiment showed that filled intervals were perceived to be longer than empty intervals, as in the previous studies, and this was clearer for 221 ms and above. For the filled intervals, the responses increased linearly as the duration lengthened, whereas for the empty intervals, the slope was steeper for durations up to 134 ms than for longer durations. The ERPs indicated negative and positive peaks related to the beginning and the end of empty intervals. In response to the filled intervals, the negative peak did not clearly appear at the end of the interval. This could reflect the overestimation of the filled time intervals, but the activities for the end of the intervals could not be observed and interpreted clearly for the 221-ms intervals. The relationship between the perceived durations of empty and filled intervals changed between 134 and 221 ms, which likely reflects a difference in the processing of empty intervals below and above this boundary.

PROMINENT INTERHEMISPHERIC FUNCTIONAL SYMMETRY IN MUSICIANS

Iballa Burunat[1,2], Elvira Brattico[3,4,5], Tuomas Puoliväli[2], Tapani Ristaniemi[2], Mikko Sams[3], Petri Toiviainen[1]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland, Finland, [2]Department of Mathematical Information Technology, University of Jyväskylä, Finland, [3]Brain & Mind Lab, Department of Biomedical Engineering and Computational Science (BECS), Aalto University School of Science, Finland, Finland, [4]Cognitive Brain Research Unit (CBRU), Institute of Behavioral Sciences, University of Helsinki, Finland, Finland, [5]Advance Magnetic Imaging (AMI) Centre, Aalto University School of Science, Finland, Finland
Time: 10:00

1. Background: Previous research has identified functional and morphometric brain differences between musicians and nonmusicians that seem to be induced by intensive musical training (Elbert et al., 1995; Schlaug et al., 1995; Ohnishi et al., 2001). However, differences in interhemispheric functional connectivity have thus far been only marginally investigated. Specifically findings on the enlarged corpus callosum in musicians, more prominently in those with an early start in musical training (Schlaug et al., 1995), and on the enhanced somatosensory left-hand finger representation in string players (Elbert et al., 1995) motivate this study.

2. Aims: We investigated the relationship between musical training and functional connectivity between homotopic (interhemispheric at homologous loci) brain areas using a naturalistic setting. We compared (a) musicians vs. nonmusicians; and (b) keyboard vs. string players. We expected to find enhanced interhemispheric functional symmetry in musicians - keyboard players in particular.

3. Methods: Participants' brain responses were recorded using functional magnetic resonance imaging (fMRI) during a naturalistic continuous listening of three music pieces of different genres. We assessed temporal coactivation of homotopic brain areas as a measure of interhemispheric functional symmetry. To tackle this, pairwise correlation of homotopic voxel time series was computed for each participant. In addition to voxelwise correlations, averages thereof per anatomical region and Brodmann area were investigated. Following this, unpaired t-tests between groups were performed. To counterbalance the neuroanatomical asymmetries, brains were transformed with a spatial mapping.

4. Results: Statistical analyses revealed significant differences between groups: interhemispheric symmetry in musicians' brain responses was more prominent over larger areas than in nonmusicians', and this was also true for keyboardists over string players. These areas were found within somatosensory, motor, visual and prefrontal cortices.

5. Conclusions: We present a novel and straightforward approach to studying inter-hemispheric functional symmetry in the brain underlying a naturalistic continuous listening. Findings evidence a positive relationship between musical training and interhemispheric connectivity. Similarly, interhemispheric connectivity is enhanced by body postural and kinematic symmetry of instrument playing. Thus results coherently build upon current knowledge on musicians' neuroanatomy.

[3B] Rhythm and Timing

KYOTO 08:30- 10:30 Tuesday 5 Aug 2014

AN EVALUATION OF MUSICIANS' INTERNAL PULSE AND RHYTHMIC SIGHT-READING

Alison Farley[1]
[1]University of Washington, USA
Time: 08:30

1. Background: Playing music at first sight involves coordination of auditory, visual, spatial and kinesthetic systems to produce an accurate and musical performance (Hayward, 2009). Accurately performing pitch and rhythm in tandem has been observed to be difficult in a sight-reading task.

2. Aims: Musicians' rhythm reading ability has a significant effect on sight-reading accuracy (Elliott, 1982). It may be conjectured that stable and consistent internal pulse is necessary to perform accurate rhythms. The purpose of this study was to assess the relationship between an individual's timekeeping ability and performance on rhythmic sight-reading tasks.

3. Method: Fifty-four wind, string or percussion instrumentalists participated in one rhythmic sight-reading and three timekeeping evaluations in two separate sessions. The sight-reading evaluation included rhythmic excerpts that increased in difficulty as participants performed the exercises out loud with a neutral spoken syllable. The timekeeping evaluations involved silently reading rhythms and keeping steady beat with and without visual notation while tapping at specified points. Absolute deviations from perfect performance were analyzed in relation to performance on the sight-reading evaluation. Participants were also asked to report any specific strategies used in performing the rhythm reading or timekeeping tasks.

4. Results: There was no significant correlation between the sight-reading evaluation and tests of internal timekeeping. A significant correlation was found between tasks involving rhythm reading and tasks focusing on timekeeping. Analysis of strategies indicated no difference in accuracy between participants who employed strategies and those who did not. Supplementary analyses were completed to determine possible reasons for the dichotomy of rhythm reading and timekeeping. Results suggested that participants had difficulty maintaining tempo rather than misreading rhythms or placing taps at incorrect points.

5. Conclusions: These data suggest that timekeeping and rhythm reading are two separate tasks. This contradicts the initial assumption that participants who perform well on rhythmic sight-reading examples would have strong time keeping ability.

SUBJECTIVE RHYTHMIZATION: A REPLICATION AND A REASSESSMENT

Rasmus Bååth[1], *Kristin Osk Ingvarsdottir*[1]
[1]*Lund University Cognitive Science, Sweden*
Time: 09:00

ABSTRACT

Subjective rhythmization (SR) is the phenomena that the sounds of a monotone metronome sequence are experienced as having different intensity and that these differences follow a regular pattern. The present study aimed to replicate and extend the two studies that have employed the original SR experimental paradigm (Bolton, 1984; Vos, 1973). The extensions included using a wider range of tempi and a large number of participants. The result of the current study was in accordance these two earlier studies. In addition to the original SR task, a novel task was administered where the participants were not explicitly told about the existence of the phenomena. The responses of the participants were in agreement with that subjective rhythmization was experienced. The result indicates that SR is a robust phenomena that can be experience even without it being primed by verbal instructions.

1. INTRODUCTION

When listening to a piece of music a common response is to move one's body with a perceived periodic pulse (Snyder and Krumhansl, 2001). That pulse is the *beat* of the corresponding piece of music, a series of subjectively isochronous (equally spaced in time) events that are felt as being pronounced or accented. The beat is established by the rhythm of the musical events and in a piece of music the beat and musical events tend to coincide. It is not necessary that every beat is marked by a musical event, however, and the perception of a beat can be sustained even if there are conflicting musical events (Large and Palmer, 2002).

It is not common that all beats in a piece of music are perceived as being equally accented (Palmer and Krumhansl, 1987) and a periodically recurring pattern of strong and weak accents is called a meter. For example, a duple meter would imply that every second beat is perceived as having a stronger accent while every third beat is perceived as having a stronger accent in the case of a triple meter. Perceiving the beat and meter of a piece of music often comes natural and it does not require the listener to actively attend to the music. It has even been shown that some form of beat induction is functional in newborn infants (Honing et al., 2009).

One perceptual phenomena that shows our tendency to experience metrical structures is subjective rhythmization (SR), a phenomena which occurs when one listens to a sequence of isochronous, identical sounds. A pattern of accents will emerge that has a metrical structure and causes the impression that there are groups of sounds. Even though the sounds are objectively identical they sound subjectively different. This phenomena was described already in the 18th century (Kirnberger, 1776) but was first investigated by Bolton (1894) who systematically played monotone metronome sequences of different tempi to a number of participants and recorded their reactions. That study was later partially replicated by Vos (1973)[1], and both studies agree on some characteristics of SR. The most common groupings participants experience are two and four, the groupings of common meters of western music. Group size and tempo interacts as participants tend to perceive smaller groupings at slower tempi and larger groupings at faster tempi, though no groupings larger than eight have been reported (Bolton, 1894; Vos, 1973). There is a limit to the range of tempi where SR can be experienced. Bolton found that participants' experience of SR ceased when the interstimuli interval (ISI) between consecutive sound onsets was above 1500 ms. After reviewing the literature Fraisse (1982) proposed that the limit was around an ISI of 1800 ms. Here is a connection to rhythm production as this limit is in the same range as when sensorimotor synchronization (e.g. finger tapping) begins to feel laborious (Repp, 2006; Bååth and Madison, 2012).

The present study aim to replicate the result of Bolton (1894) and Vos (1973) using a wide range of tempi and a larger number of participants than in the earlier studies. Such a replication is presented in Experiment 1. A second aim was to verify the existence of the SR phenomena. In previous studies SR has been investigated by explicitly asking what groupings participants experience when listening to a metronome sequence. By explicitly asking participants about grouping it is possible that participants get primed to experience SR. In Experiment 2 we used a novel task where the response of a participant depends on whether he or she experiences SR but where SR is not mentioned in the task instructions.

2. EXPERIMENT 1

2.1. Method

Participants were recruited, and the experiment administered, using the on-line service Amazon Mechanical Turk (Buhrmester et al., 2011). Out of the 132 participants 111 reported having experience playing a musical instrument. The task instruction given to the participants were as follows:

"This task requires your full attention. Below are six sound sequences of clicks. You should listen to each sound sequence and rate if you feel any grouping or subdivision of the clicks, however weak or subtle. For example, if you hear "TICK-tick-TICK-tick-TICK-tick" that would be groups of two, while if you hear "TICK-tick-tick-tick-TICK-tick-tick-tick" that would be groups of four. This task **is not** about whether there are groups in the sequences, it is about if you **feel** any grouping. Now listen to and rate the sequences one at a time in the order they are displayed."

The participants were then given six monotone metronome sequences with ISIs of 200 ms, 300 ms, 400 ms, 600 ms, 800 ms and 1500 ms in a randomized order. Each sequence was 15 s long and consisted of 10 ms long, 440 Hz sine wave sounds. After listening to a sequence the participant indicated what grouping he or she felt by using a list with the alternatives "no group" and "groups of 2" up to "groups of 8".

2.2. Result

The result replicated many of the findings of Bolton (1894) and Vos (1973) and Figure 1 shows the reported experienced grouping from six participants. In general participants reported larger groupings at faster tempi, as can be seen in participant C or F. Some participants were consistent, like participants A and B, while some were less consistent, like participant C and E. There was a tendency that some participants (for example, participant D) answered "no grouping" on the faster tempi (200 ms and 300 ms).

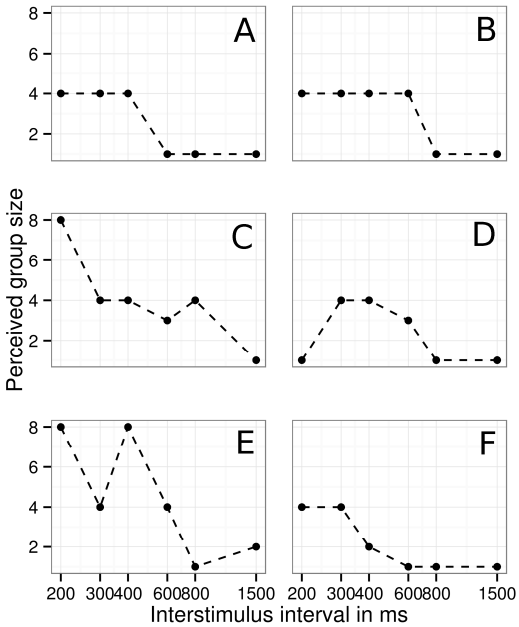


Figure 1: The reported perceived group size for six of the 132 participants in Experiment 1. Note that a group size of one corresponds to the participant having reported "no grouping".

The most reported groupings were two, four and eight with five and seven being rarely reported at all. Table 1 show how often the participants reported each possible grouping. Figure 2 show for each ISI level the proportion of participants that reported each grouping. The slower limit of SR was estimated to an ISI of 1500 by Bolton. In the current study no such sharp limit was found, however, a majority of the participants (81 %) reported experiencing no grouping at an ISI of 1500 ms.

Table 1: Summary of the reported groupings in Experiment 1.

Grouping	Peak ISI	% of responses
1 (No grouping)	1500	34 %
2	600	23 %
3	300	7 %
4	300	27 %
5	200	< 1 %
6	300	1 %
7	200	< 1 %
8	200	6 %

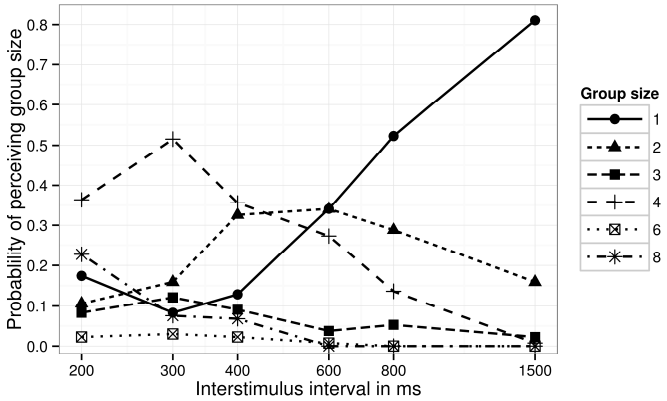


Figure 2: Percentage of reported groupings as a function of ISI. A group size of 1 corresponds to "no grouping".

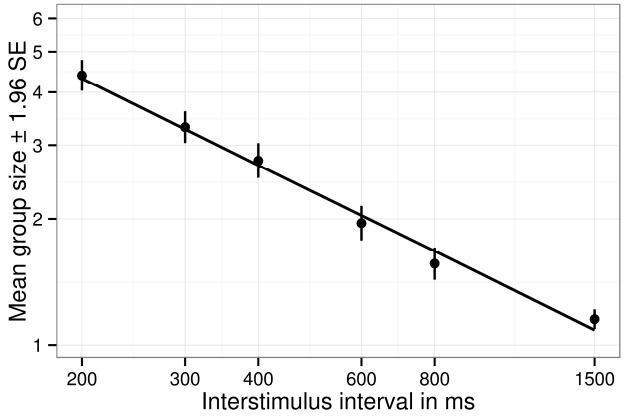


Figure 3: Log-log plot of mean group size as a function of ISI. The line show the best fitting regression line.

Figure 3 shows the mean group size reported for each ISI level with both axes being on the log scale. This relation appears highly linear, a result not previously reported in the literature. A linear regression with $\log_2$ group size as the dependent variable and and $\log_2$ ISI level as the independent variable gave an intercept of 8.1 (95% bootstrap3 confidence interval [5.8, 10.5]), a slope of -0.77 (95% CI [-1.0, -0.53]) and an R^2 of 0.50 (95% CI [0.23, 0.77]).

3. EXPERIMENT 2

3.1. Method

The purpose of Experiment 2 was to investigate whether the phenomena of SR would influence participants responses even though SR was not mentioned in the instructions nor suggested in any way. Amazon Mechanical Turk was again used to recruit and administer the task to 120 participants, 60 in each of two conditions, where the only difference between the conditions were whether the following task instructions used the word *second* or *fourth*:

"In this task we are interested in if it is possible to feel very small differences in loudness. Below are 14 click sequences, in some of them all clicks are equally loud and in some of them every [*second, fourth*] click is a little bit louder. The difference in loudness will be

3 The bootstrap confidence intervals (CI) were calculated using 10,000 resamples.

very small. Listen to the sequences, in the order they are given, and for each sequence try to feel if the clicks are equally loud or if every [*second, fourth*] click is louder."

The 14 click sequences were the same as in Experiment 1, with the addition of a 2000 ms ISI sequence, each given twice in a randomized order. That is, despite the task instructions, all clicks were actually equally loud. After having listened to each sequence the participant was asked whether they perceived a difference in loudness or not. If the participant, without knowing it, experienced SR when listening to the sound sequence we would expect her to be more likely to report a difference in loudness. If a participant was given the *second*-instructions, to listen for a difference on every second click, we hypothesized that she or he would direct attention towards SR with a grouping of two and therefore be most likely to report a difference at an ISI of around 600 ms (cf. Figure 2). Similarly, if a participant was given the *fourth*-instructions and was listening for a difference on every fourth click, we would expect him or her to be most likely to report a difference around an ISI of 300 ms.

3.2. Result

There was a tendency for the participants given the *second*-instructions to report hearing a difference around an ISI of 600 ms while the participants given the *fourth*-instructions were more likely to report a grouping at the ISIs of 300 ms and 400 ms. Figure 4 show the probability of reporting hearing a difference for each ISI level, where there is a clear peak around 600 ms for the *second*-condition and around 300 ms and 400 ms for the *fourth*-condition. At the ISI levels of 1500 ms and 2000 ms, where the participants in Experiment 1 largely reported hearing no grouping, there is an increase in reporting hearing a difference for the *fourth*-condition. A reason for this could be that at such slow tempi some participants find it difficult to compare every fourth click and therefore approach the chance level of 50%.

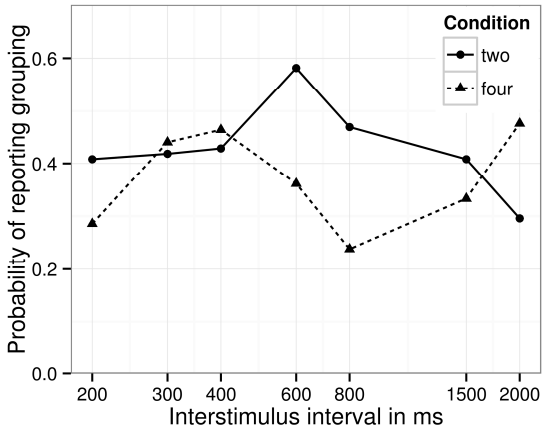


Figure 3: Log-log plot of mean group size as a function of ISI. The line show the best fitting regression line.

4. CONCLUSION

Experiment 1 replicated the main findings of Bolton (1894) and Vos (1973):

- Subjective Rhythmization is a robust phenomena that seems to be experienced by most participants.
- The reported experienced grouping is most often two, four or eight, common meters of western music.
- What grouping that is reported is highly dependent on the tempo with larger groupings being reported at faster tempi.
- Most participants do not reporting hearing a grouping when the ISI is as slow as 1500 ms.

Experiment 2 confirmed the robustness of SR and showed that SR influence participants' responses in a task that is quite different from the original SR task.

It should be noted that as both experiments used Amazon Mechanical Turk, both experiments were administered on-line without the usual control of a perceptual experiment. This can be viewed both as a weakness and as a strength, a weakness because there was no control over what environment the participants were in when doing the experiment, a strength because despite the lack of control the result is well in agreement with the earlier studies of Bolton (1894) and Vos (1973).

5. ACKNOWLEDGEMENT

This research was supported by the Linnaeus environment Thinking in Time: Cognition, Communication and Learning, Swedish Research Council grant number 349-2007-8695.

6. REFERENCES

Bolton, T. (1894). Rhythm. The American Journal of Psychology. *Retrieved from http://www.jstor.org/stable/1410948*

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspectives on Psychological Science, *6(1), 3–5. doi:10.1177/1745691610393980*

Bååth, R., & Madison, G. (2012). The subjective difficulty of tapping to a slow beat. In Proceedings of the 12th International Conference on Music Perception and Cognition *(pp. 82–85). Retrieved from http://www.sumsar.net/papers/ ICMPC_2012_rasmus_baath.pdf*

Honing, H., Ladinig, O., Háden, G. P., & Winkler, I. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials. Annals of the New York Academy of Sciences, *1169(Illc), 93–6. doi:10.1111/j.1749-6632.2009.04761.x*

Kirnberger, J. P. (1776). Die Kunst des reinen Satzes in der Musik: Aus sicheren Grundsätzen hergeleitet und mit deutlichen Beyspielen erläutert. *Decker.*

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. Cognitive Science, *26(1), 1–37. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S036402130100057X*

Van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. Journal of New Music Research, *28(1), 43–66. doi:10.1076/jnmr.28.1.43.3122*

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. Journal of Experimental Psychology: Human Perception and Performance, *13(1), 116–26. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2951485*

Snyder, J., & Krumhansl, C. L. (2001). Tapping to Ragtime: Cues to Pulse Finding. Music Perception, *18(4), 455–489. doi:10.1525/mp.2001.18.4.455*

Vos, P. (1973). Waarneming van metrische toonreeksen. *Nijmegen: Stichting Studentenpers.*

TAPPING THROUGH LONGER PERIODS OF SILENCE IMPROVES PERCEIVED TIMING

Fiona Manning[1], Michael Schutz[1]
[1]McMaster University, Canada
Time: 09:30

1. Background: Musical situations comprise both rhythmic sequences of notes and silences that separate these sequences. While changes in timing occur frequently in musical passages, we know that various temporal contexts can influence perceived timing in different ways (McAuley & Jones, 2003: Repp, 2002).

2. Aims: Although we know that moving with the beat can improve perceived timing (Manning & Schutz, 2013), the role of movement through different durations of silence of remains unclear. Here, we manipulated the duration of silence embedded in a sequence, to examine the role of movement with and without an external pacing stimulus.

3. Method: Participants listened to an isochronous sequence of beats and identified the timing of a final probe tone. They either tapped along with the sequence of beats, or listened while remaining still. In one experiment the probe tone occurred immediately after the sequence (i.e., without a period of silence). In additional experiments the probe tone was separated from the sequence using a silence equivalent to the duration of one, two or three beats. For all experiments, the probe tone fell on a metrically strong beat should participants be counting in duple meter.

4. Results: In the experiment that contained no silence, participants performed the deviation detection task better when listening alone, compared to when they tapped along with the sequence. In the experiment that contained one beat of silence, participants performed better on the task when moving along in the late probe tone offsets but worse when moving alone in the early offsets. When the experiment contained two or more beats of silence, participants perform better on the task only when moving along compared to when listening alone.

5. Conclusions:
Overall, these experiments demonstrate that the longer the duration of silence, the greater the benefit of moving with the beat. This set of studies shows that while silence embedded in rhythmic sequences increases demands on memory and beat tracking, movement can promote beat tracking when auditory information is absent.

OPTIMISING A TEST OF THE ABILITY TO TAP IN TIME WITH MUSICAL BEAT

Jason Musil[1], Daniel Müllensiefen[1]
[1]Goldsmiths, University of London, United Kingdom
Time: 10:00

1. Background: Humans commonly move to music, however there are meaningful individual differences in the accuracy with which people synchronise actions (entrain) to musical beat (Iversen & Patel, 2008). Previous entrainment research has focused mostly on underlying cognitive processes, or group differences in ability, but not on population variability. However, investigating entrainment within an individual differences framework will reveal how performance differences on this "common" ability relate to other musical abilities, training, and engagement.

2. Aims: The first aim of this study was to develop a short, valid, and reliable test of tapping synchronisation ability using real musical excerpts. A second aim was to compare synchronisation ability with beat perception accuracy, and with other musical listening skills and self-reported musical training and engagement.

3. Method: 297 participants were recorded tapping to the beat of 37 short music excerpts with various tempos, meters, and genres, and item response analyses were performed on the data. Circular mean offsets and standard deviations were computed against beat locations from averaged high-agreement performances by two professional drummers, and quantile-transformed to give an accuracy index. A second dataset is being collected to validate the resulting beat synchronisation test against other tasks of the Goldsmiths Musical Sophistication Index battery (Müllensiefen et al., 2014) as well as the 'rhythm' and 'speed' tasks from the PROMS (Law & Zentner, 2012).

4. Results: The item response analyses yielded a 22-item beat synchronisation test battery that takes about 7 minutes to complete and requires no formal musical knowledge. Results from a current validation study will establish convergent and divergent validity, internal consistency, and test-retest reliability of the new test battery.

5. Conclusions: We present a simple method of assessing tapping synchronisation ability, using test items that fit a rigorous item response measurement model. Circular mean offset and variability measures are arguably analogous to phase and tempo errors in a beat alignment perception task, which will allow the relationship between production and detection of phase and tempo errors to be analysed in a subsequent step, thus linking cognitive and individual differences research into this ability. The tapping test will be made freely available to the research community.

[3C] Musical Structure

MONTREAL 08:30- 10:30 Tuesday 5 Aug 2014

LISTENERS' INFORMAL VOCABULARY FOR STRUCTURE, EXPRESSION, EMOTIONS AND ASSOCIATIONS IN PIANO MUSIC

Marlies Bodinger[1], Erica Bisesi[1], Richard Parncutt[1]
[1]University of Graz, Austria
Time: 08:30

ABSTRACT

How do different people use words to describe music? Where are their differences and similarities? And what exactly is it what they describe? Are they talking about music structure, emotions, free associations, or other things? This qualitative study aims at answering these questions in the context of grounded theory. We are interviewing 28 participants belonging to 4 categories of expertise: musicologists, expert musicians, classical music lovers and non-experts. All participants listened to 8 short music excerpts in different and contrasting classical music styles, selected in a previous study whose aim was to provide a set of pieces covering the whole spectrum of categories for emotion as predicted by the Hevner Adjective Clock. A preliminary analysis of data in MAXQDA suggested a relationship between participants' expertise and their descriptions (expert musicians focused more on emotion, music lovers on free associations), and a co-occurrence of categories related to the musical surface (melodic/harmonic expectation, rhythm, tempo, dynamics), emotional valence/arousal, motion/commotion, and contrasting free associations (static/dynamic, living/non-living things).

1. BACKGROUND

Literature about music interpretation is mostly written by performers (e.g. Barenboim and Said, 2004; Brendel, 1991; 1997; delle Vigne, 2006), but does not necessarily reflects the way musicians think in their everyday life. There exist many approaches of doing research in music. One traditional approach is historical: a researcher may ask about what happened at particular times, what did it mean, and what does it mean today. On the other side, a very common approach is experimental: how can a certain result be measured, analyzed and coded? One can address the same questions from also other perspectives, e.g. focusing on philosophical or sociological aspects, or just looking at a description and/or classification of data.

All of these approaches may have quantitative and qualitative aspects. Generally, quantitative research enables a researcher to control some variables when manipulating others, but one can sacrifice reality for the sake of control. On the other side, qualitative research may provide very copious descriptions and details, but it suffers the restriction that conclusions may have a limited generalization - as subjectivity generally leads to procedural problems due to unavoidable researchers' bias. Nevertheless, a qualitative approach may offer a diverse number of advantages. By using subjective information, researchers may get an in-depth examination of phenomena, addressing complex questions that can be impossible with quantitative methods. In the past several years, there has been an increasingly interest in qualitative methods in music research, mostly focusing on music education, music sociology and ethnomusicology (Bresler, 2008; Radocy, 1994), as indicated by the substantial increasing of qualitative reports published in music conferences and journals (Lane, 2011).

On the other side, research in experimental aesthetics has been greatly aided by the advent of the Hevner adjective list (Hevner, 1935; 1936). After its formulation, the Hevner checklist - consisting of clusters of words grouped by similarity of meaning, arranged in a circle so that adjacent clusters are more similar in meaning than more distant clusters - has been progressively revised by improving its consistency across and within each cluster (Farnsworth, 1954; 1969), and updating the checklist presenting a wider spread of emotional ranges (Schubert, 2003).

We addressed the question how listeners' use informal vocabularies for describing expression in piano music for the first time in 2010, and found that words fell into three categories: musical structure/expression, emotion, and free association (Bisesi and Parncutt, 2010). Subcategories included immanent/performed attributes of the sound (tempo, dynamics, pitch and timbre), accentuation, character and meaning, basic/complex emotions and static/dynamic or living/non-living associations.

2. AIMS

The current study investigates the relationship between music structure, emotions and free associations in more detail. Our goal is to modify the Hevner categorical model of emotions and/or find an alternative way of clustering the narrative data gathered from the interviews. In order to reveal new forms of clustered information concerning listeners' informal vocabulary for classic piano music, we are using a narrative method for the process of data gathering (Zentner and Eerola, 2011).

3. METHODS

3.1. Participants

28 participants (7 musicologists, 7 expert musicians, 7 classical music lovers and 7 non-experts) were recruited by email.

3.2. Materials

Stimuli were short excerpts from 8 pieces of different and contrasting classical music styles, covering a broad spectrum of emotions and piano repertoire from baroque to modern music.

The question of what pieces are to be used in order to cover the broadest spectrum of emotions and feelings (as predicted by the Hevner Adjective Clock, cfr. Hevner, 1935; 1936) has been addressed in a pilot research conducted by Florian Eckl and Erica Bisesi in 2013. Starting with a selection of 35 pieces (as done subjectively by the researchers themselves by looking at all classical piano styles from the 18th to the 20th century repertoire, as well as at contrasting emotions), we selected the 8 pieces which resulted to be the more appropriate to represent each of the 8 Hevner categories of emotions. To identify these 8 pieces, 21 participants (7 musicians, 7 amateurs and 7 non-musicians) were asked to associate to each of the original 35 pieces the 5 adjectives that they thought best represented each piece, out of a set of 70 words (the 66 Hevner adjectives, plus 4 additional adjectives provided by Schubert (2003)), and to rate each adjective with a score from 1 to 5. All the adjectives were presented to participants in their mother language (German).



Figure 1: English translation for the computer interface (designed by means of the Psychopy open-source application), including all the words presented to participants in the pilot study (originally in German).

Figures 1 shows the computer interface (developed by means of the Psychopy open-source application) used in this pilot study. Best-rated pieces are shown in Figure 2. They are short excerpts (lasting 30-60 sec) from: J. S. BACH, Fugue No.22 in B flat minor BWV 867; F. LISZT, La Lugubre Gondola I (S.200); R. SCHUMANN, Davidsbündlertanz Op.6 No. 14; L. van BEETHOVEN, Sonata Op.31 No.2 (2nd Mvt.); W.A. MOZART, Allegro K deest; D. SCARLATTI, Sonata K159 in C major; K. SZYMANOWSKY, Prelude Op.1 No.5 in D minor; F. CHOPIN, Sonata in B minor Op.58 (1st Mvt.). These pieces have been selected as stimuli for the study described in the following of this paper. In both stages of the procedure described below, the dynamics of all the pieces was normalized by means of the open-source Audacity music editor.

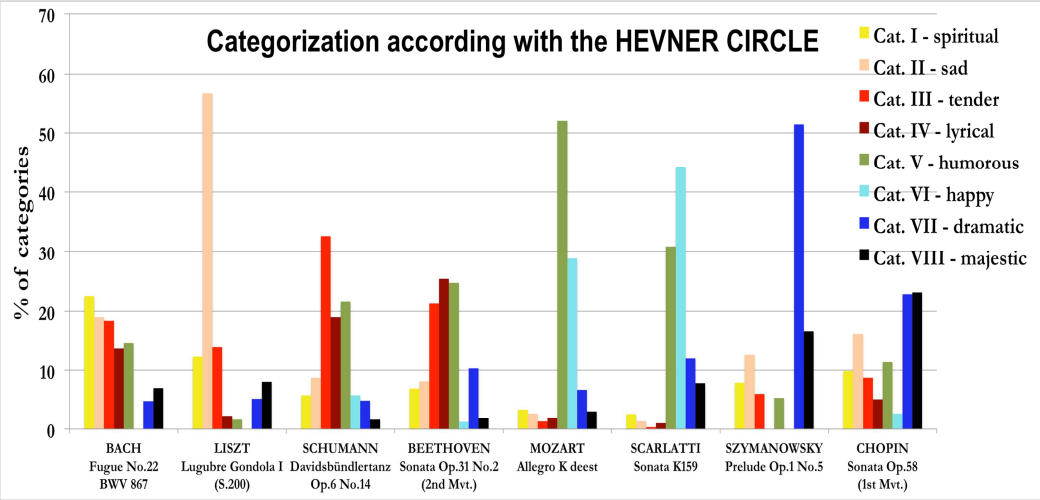


Figure 2: Categorization of best-rated pieces as resulting from the pilot study.

3.3. Procedure

For this research design, we adopted the grounded theory approach (Glaser and Strauss, 1967) - a qualitative research methodology in the social sciences involving the generation of theory from data. In grounded theory, rather than beginning with a hypothesis, the first step is data collection. Further, the key points are coded and grouped into concepts and categories, which are the basis for the creation of a theory. In this way, it's possible to find out new ways of building typologies, clusters and categorizations out of the gathered information.

Data collection was carried out by means of *guideline-based, narrative interviews*. The first set of questions included the approach to music in everyday life (i.e. when and which style of music people listen to, and whether they play an instrument or not) (Table 1).

Table 1: Interview-guideline, part 1

Questions	Topics
In which situations do you listen to music?	Motivations and goals Feelings What kind of music (genre, style) Active or passive listening Sources
What style/genre?	Instrumental or with lyrics
What is really evoking an emotion? (Differentiation of felt vs. perceived emotion)	Melody, lyrics, other aspects in the structure
Do/did you play an instrument?	Which one? How long?

After having answered to this first set of questions, participants were asked to listen to the first piece on headphones and to freely talk about *what they heard, felt and thought during the process of listening* (according to a second interview-guideline, see Table 2). This procedure was then repeated for all the 8 pieces listed above. Each participant listened to the pieces in a different random order.

Table 2: Interview-guideline, part 2.

Questions	Topics
Ways of describing the piece	Structure, emotions, free associations Positive/negative experience, intensity of the experience
How did the piece affect you?	Because of structure (melody, harmony, timbre, rhythm, tempo, melody, phrasing, accentuation) Because of emotion (felt/perceived, physical involvement)
Did you "see" something?	Because of associations (living/non-living things, colors, people, situations, fantasy, memories, …) Other (technical aspects, role of the performer)
Did you know the piece?	How, how long, how often have you listened to it? Have you never performed it?
Would you also listen to this piece in private?	In which situations?

4. PRELIMINARY RESULTS AND DISCUSSION

Our pilot study revealed some aspects of weakness in the Hevner categorical model. For instance, we found that only 5 out of 8 Hevner categories of adjectives are well represented, despite the relatively high number of pieces used to conduct this study (see Figure 2). Moreover, words frequently overlap (an issue also pointed out by most of the participants, and consistent with previous studies, cfr. Farnsworth, 1954; 1959; Schubert, 2003). All these findings strengthen the evidence for the need of an accurate revision of the model. At the current stage, we have conducted interviews, coded and preliminarily analyzed data for 16 participants. The data processing is carried out by means of the computer software for qualitative data analysis MAXQDA 11. After transcribing all the interviews, the transcripts are coded in order to highlight the main topics emerging from the interviews.

In general, participants talked quite a lot about their personal approach, appreciation and knowledge of the music. Depending on the piece, they mentioned structure, emotion, and free associations in greater or lesser extent. As expected, there exists a strong relationship between participants' expertise and their descriptions: expert musicians focused more on emotion, music lovers on free associations.

We also found a co-occurrence of categories related to the musical surface (melodic/harmonic expectation, rhythm, tempo, dynamics), emotional valence/arousal and motion/commotion, and between some aspects in the music surface (mostly tempo and dynamics) and contrasting free associations (static/dynamic, living/non-living things). Further details (e.g., formulation of typologies, categories and clusters of participant) are in preparation and will be discussed at the conference.

Results will be applied to several research contexts. For instance, the revised Hevner model and its extension to free associations will influence quantitative studies aimed at mapping music expression into multidimensional spaces of emotions and free associations - such as the formulation of a model of immanent emotion as a linear combination of simple predictors for valence and arousal (strength of tonal center, dissonance and average pitch variation; cfr. Bisesi et al., 2014), and the design of a new version of pDM (a system for real-time expressive control of music performance; cfr. Friberg et al., 2006).

5. ACKNOWLEDGEMENTS

This research is supported by the FWF Stand-Alone Project P 24336-G21 "Expression, emotion and imagery in music performance". We kindly thank Florian Eckl (KFU, Graz) for his support in the preliminary stage of this study.

6. REFERENCES

Barenboim, D., & Said, E. V. (2004). Parallels and paradoxes: Explorations in music and society (Reprinted edition). New York: Vintage Book.

Bisesi, E., & Parncutt, R. (2010). The informal vocabulary of professional musicians for describing expression and interpretation. Paper presented at ICMPC11, Seattle, WA, USA, 2010.

Bisesi, E., Parncutt, R., Caron, S., & Traube, C. (2014). The immanent emotion of a musical score: An exploratory study. Abstract accepted for spoken presentation at ICMPC-APSCOM 2014, Seoul, Korea, 2014.

Brendel, A. (1991). Musical thoughts and afterthoughts (Reissue edition). New York: Noonday Press.

Brendel, A. (1997). Paradosso dell'interprete. Pensieri e riflessioni sulla musica. Firenze: Passigli.

Bresler, L. (2008). Music in qualitative research. In Lisa M. Given and Gale Group (Eds.), The SAGE Encyclopedia of Qualitative Research Methods. *Los Angeles, Calif: SAGE Publications Inc.*

delle Vigne, A. (2006). Viaggio nell'intimo di un pianista. Bari: Florestano.

Farnsworth, P. R. (1954). A study of the Hevner adjective list. Journal of Aesthetics and Art Criticism, 13, *97-103.*

Farnsworth, P.R. (1969). The social psychology of music (2nd Ed.). Ames, I.A.: State University Press.

Friberg, A. (2006). pDM: An expressive sequencer with real-time control of the KTH music performance rules. Computer Music Journal, 30 (1), *37-48.*

Lane, J. (2011). A descriptive analysis of qualitative research published in two eminent music education research journals. Bulletin of the Council for Research in Music Education, 188, *65-76.*

Glaser, B. G., & Strauss, A. L. (1967). The discovery of Grounded Theory. Strategies for qualitative research. Chicago: Aldine.

Hevner, K. (1935). The affective character of the major and minor modes in music. American Journal of Psychology, 47, *103-118.*

Hevner, K. (1936). Experimental studies of the elements of expression in music. American Journal of Psychology, 48, *246-268.*

Radocy, R. E. (1994). Qualitative research in music education: Why the fuss? Bulletin of the Council for Research in Music Education, 122, *94-103.*

Schubert, E. (2003). Update of the Hevner adjective checklist. Perceptual and Motor Skills, 96, *1117-1122.*

Zentner, M., & Eerola, T. (2011). Self-report measures and models. In P. Juslin (Ed.), Handbook of music and emotion: theory, research, applications *(pp. 187-221). Oxford: Oxford University Press.*

VERBATIM REPETITION AND LISTENING MUSICALLY

Elizabeth Margulis[1]
[1]*University of Arkansas, USA*
Time: 09:00

1. Background: The topics of musical similarity and variation have received considerable attention both from psychology (Deliège, 2007) and musicology (Sisman, 1993; Zbikowski, 2002). Less well studied has been its correlate, exact repetition. A recent book (Margulis, 2013) attempts to remedy that oversight, synthesizing work from a number of fields to outline a theory about the psychological processes

underlying the enjoyment of musical repetition. An interesting borderline case between the poles of repetition and variation is performer nuance. Even when the notation is identical among different iterations of a principal theme, performers often change expressive inflections, altering dynamics and microtiming on each rendition.

2. Aims: This study aims to investigate the perceptual consequences of verbatim repetition (where every sound is exactly replicated) in comparison with varied repetition (where performer nuance subtly alters the sounds between restatements).

3. Method: 31 participants without formal musical training listened to eighteenth-century rondos featuring several repetitions of a main theme. They heard some of the rondos in Modified form, where the theme was played slightly differently on each occurrence, and some in Verbatim form, where the recording had been digitally altered so that each statement of the theme was exactly the same. Participants answered questions about their experience after each excerpt and then, in the second phase of the experiment, performed a recognition memory task.

4. Results: Excerpts heard in the Verbatim condition were rated as more repetitive by listeners, but were also judged to have made listeners more likely to move, sing, or tap along to the music. In addition, when presented with a short snippet of the theme, listeners in the Verbatim condition were able to play through the theme's continuation in their head for a longer duration than listeners in the Modified condition. Repetition did not improve recognition memory for snippets from the original version of the theme.

5. Conclusions: Taken together, these results suggest that verbatim repetition enhances a listener's sense of subjective identification and participation with the musical trajectory of a piece, lending new insight into the way that musical repetition works its well-known effects on enjoyment and aesthetic experience.

FREQUENCY AS DETERMINANT OF FUNCTION AND STRUCTURE: A USAGE-BASED APPROACH TO JOÃO GILBERTO'S GUITAR ACCOMPANIMENT IN BOSSA NOVA

Bruno Alcalde[1]
[1]*Northwestern University, USA*
Time: 09:30

A prominent characteristic of Brazilian Bossa Nova style is its rhythmic guitar accompaniment, which has in João Gilberto a chief reference. Working with a restricted number of basic rhythmic patterns and variations, Gilberto reserves less frequent patterns for specific points in the pieces. This paper investigates these choices of rhythmic configuration from a usage-based perspective. Based on an initial corpus of nine pieces from the first three records of the singer/guitarist's discography, the thesis is that the use of less frequent patterns in the right-hand rhythm serves to emphasize important features of the musical structure. Furthermore, in investigating this small corpus one can begin to establish a "grammar" of João Gilberto's guitar playing (and of Bossa Nova in general) that is expressed through accumulation of information in the development of the style. In this repertory, rhythm serves as one of the main markers of style, being a crucial aspect for its recognition. The transcription process resulted in 622 measures of rhythmic patterns, making a total of 39 different 1-measure patterns that were categorized according to their frequency in the corpus. Four interrelated patterns (two basic patterns and their variations with tied first beats) represent 75% of the transcribed rhythms. Each song was analyzed in a "frequency of pattern vs. measure" graph. This afforded a clear visualization of the frequency groups interacting with several levels of the structure: section, phrases, harmonic progressions, and melodic activity. In these graphs, high peaks will show uncommon patterns in the corpus that serve as markers of structural features. Overall, the analyses show six different aspects that are emphasized by the use of the less frequent patterns throughout the songs: (1) articulation of form; (2) dominant chords; (3) secondary dominant chords; (4) cadential regions; (5) highest pitch of the melody; and (6) two chords in a single measure. In order to explore the stability of theses stylistic features over time, an additional analysis was made comparing two recordings of the same song more than 40 years apart. The comparison shows that most of the peaks and valleys profiles were kept, but some others were reversed: what was peak became a valley and vice-versa. This suggests that an additional important aspect is the contrast of frequencies as marker of structure. In this paper, I contextualize and explain these ideas, as well as provide analytical examples of the six categories mentioned above using João Gilberto's recordings.

AN INTEGRATIVE COMPUTATIONAL MODELLING OF MUSIC STRUCTURE APPREHENSION

Olivier Lartillot[1]
[1]*Aalborg University, Denmark*
Time: 10:00

ABSTRACT

An objectivization of music analysis requires a detailed formalization of the underlying principles and methods. The formalization of the most elementary structural processes is hindered by the complexity of music, both in terms of profusions of entities (such as notes) and of tight interactions between a large number of dimensions. Computational modeling would enable systematic and exhaustive tests on sizeable pieces of music, yet current researches cover particular musical dimensions with limited success. The aim of this research is to conceive a computational modeling of music analysis encompassing the main core components and focusing on their interdependencies. The system should be as simple as possible, while producing relevant structural analyses on a large variety of music. This paper describes the general principles of a computational framework for music analysis currently under construction. The complex system of music structure is described as an interaction between modules focusing on formal operations that are conceived as general as possible. Each module addresses a core aspect of music analysis and offers some innovative breakthrough compared to the state of the art. In order to overcome the limitations of local segmentation, we propose an alternative paradigm based on hierarchical local grouping. Ornamentation

reduction, and more generally the inference of deeper structures from the musical surface is carried out by selecting key elements from each local grouping structure, offering hence an implementation of "Time-Span Reduction". Parallelism, i.e. sequential repetition, considered here as an essential aspect of music analysis, is applied to the search for motives, for mode-related patterns as well as metrical analysis. A new approach for modal analysis is based on a comparison of the local context (defined by the current and recent notes, and taking into account ornaments reduction) with all possible modes and key scales. Structural information inferred in each module is fed back into the other modules. Besides the practical application to automated music analysis, the computational model, by virtue of its generality, extensiveness and operationality, is suggested as a blueprint for the establishment of cognitively validated model of music structure apprehension. Available as a Matlab module, it can be used for practical musicological uses.

1. MOTIVATIONS

Music analysis is aimed not only at identifying known theoretical schemes on particular pieces of music, but also at revealing the particularities and richness of each individual composition or improvisation. This gets particularly challenging when considering such ideal in a radical perspective of exhaustive characterization of the composer's or musician's structural intentions (Reti, 1951). Indeed, music is extremely complex, both quantitatively (as a long and dense flux of notes intricately interconnected) and qualitatively (music is described along a complex taxonomy of musical dimensions). For all the creative interests of subjective and hermeneutic analyses, a more rigorous procedure based on systematizing explicitly formalized rules allows a better assessment and control of the knowledge thus inferred, no more relying on a given musicologist's authority, but rather on a well-defined approach. But of course, formalizing such systematic procedure is far from easy, and applying that model rigorously is not simple as well4. Lerdahl and Jackendoff (1983)'s *General Theory of Tonal Music* (GTTM) offers one of the most comprehensive models for music analysis. The authors acknowledged however that some of the most challenging aspects of music (such as parallelism) cannot be easily systematized in their approach, and in order to cope for the high complexity of music, they had to rely on a strong reductionist model.

Computational representation and automation offer a precious help to this endeavor, enabling to carry out exhaustive and rigorous execution of systematized models on corpora of music of any size. The implacable way the computer reveals the validity of systematized models is particular telling: it emphasizes in particular the difficulty of designing a model that can produce results that are at least verisimilar. This shows in particular that, in opposition to Ruwet (1951) or Nattiez (1987), there is no simple way to extract structures that would exist immanently, but on the contrary that these structures need to be constructed according to particular heuristics. Because the analysis cannot rely on the composer's or musician's explanations5, these heuristics need to rely, at least partially, on cultural knowledge and perceptual principles6.

Significant research has been carried out in computational music analysis, offering interesting tools that however have not succeeded yet to automate most of the core facets of traditional music analysis' praxis. One major hypothesis underlying the research presented in this paper is that music is a complex phenomenon, composed of a network of components that interact tightly. This can be understood in a purely musicological point of view, as an interconnection of musical dimensions, but also in a cognitive perspective. As such, in the same way that music analysis cannot focus on one single dimension without considering the other dimensions that interact, computational modeling of music analysis need to take into account this integrated network altogether.

This paper describes the general principles of a computational model currently under construction that follows these principles. The complex system of music structure is described as an interaction between modules focusing on formal operations that are conceived as general as possible.

2. LOCAL GROUPING

There has been significant research around the concept of *local segmentation*, studying the emergence of structure related to the mere variability in the succession of musical parameters. These researches, notably by Tenney and Polansky (1980) or by Cambouropoulos (2006), focus on the analysis of monodies, and model this structural phenomenon as a segmentation of the monody, which cuts the temporal span at particular instants, resulting into a linear succession of segments. In these approaches the heuristics for segmentation is based on a mixture of several constraints related to what happens *both before and after* each candidate segmentation point, which leads to approximate and incomplete segmentation results (Lartillot, Yazıcı & Mungan, 2013). We presented instead a simpler approach focused only on what happens *before* each candidate segmentation: this enables to reveal a more complete set of segmentation points, to indicate more precisely the temporal locations of the segmentation points, and could also reveal a segmentation hierarchy at multiple structural levels.

We introduce a new formulation of our proposed approach that reveals a much clearer structural description and that can be explained with simple principles (Lartillot & Ayari, 2014). The approach focuses on *grouping* instead of *segmentation*. In other words, what needs to be characterized are not the segments between notes, but instead the groups of notes that are progressively constructed. This clustering mechanism operates differently on time and pitch domains.

4 For instance, Ruwet (1951) did neither actually apply the model he introduced, nor test its validity.
[5] The author of such a complex creative process would not easily describe its underlying mechanisms in all its details anyway.
[6] Music theory could be related to these two aspects.

In the time domain, local grouping can aggregate in a purely hierarchical fashion. Each group is characterized by the maximal temporal distance between successive notes7. Smaller groups are related to short temporal distances while larger group that contains other groups are related to longer temporal distances. For instance, in Figure 3, the piece starts with four eighth-notes followed by a fourth-note. These first 5 notes (with successive pitches C,C,C,D,E) form a local group related to the rhythmic value of eighth-note. This 5-note local group is followed by an isolated fourth-note (D) and a second 5-note local group (C,E,D,D,C). Since these 11 notes end with a half-note, they form altogether a higher-level local group related to the rhythmic value of fourth-note.

In the pitch domain, this clustering is related to the idea of voice streaming, i.e. of tracking monodies within polyphonies. In that sense, the mere comparison of inter-pitch intervals between successive notes, as done in previous work in segmentation, does not suffice, because monodic voices can form between notes that are not directly successive in the music surface. The formalization of voice streaming is under study.

3. ORNAMENT REDUCTION

Previous computational attempts to model processes related to melodic reduction in music (Gilbert & Conklin, 2007; Marsden, 2010) do not formalize concrete musicological or cognitive conditions for reduction. We present a set of rules founding the detection of ornaments, based on local grouping.

3.1. Local Group Head

By definition, a time-based local group terminates with a note that is followed by a duration (before the next note) that is significantly longer than the temporal distance between notes within the group. As such, the local group can be perceived as a phrase that terminates with a concluding note that has a more structural importance. This hypothesis might not be always valid, in particular in the presence of particular accentuations at particular notes within the group. But in more general case, it seems to offer some general interest. Following this observation, we propose to formalize this hierarchy of notes in local groups by associating with each local group a main note, or "head" to follow the GTTM's *Time-Span Reduction* terminology, which would in the simple case be the last note of the group, as circled in red in Figures 3 and 4. The other notes would be considered as "subordinate events" or – why not – as the "tail" of the group. For instance, in Figure 3, the first 5-note local group (C,C,C,D,E) terminates with the longer note E, becoming the head of the group.

When subordinate events in a local group have same pitch than the final note of the group, they all form a single note – a "*meta-note*" (Lartillot & Ayari, 2012) –, which will become the actual head of the group. The subordinate events of the groups can be considered as forming an ornament – such as a *cambiata* or a *trill* – of the group's head. For instance, in Figure 3, the second 5-note local group (C,E,D,D,C) ends with pitch C, which already appears at the beginning of the group, thus the two occurrences of pitch C form a meta-note, which is highlighted by a red rectangle. The local group develops a *cambiata* around that main note C.

Meta-note can develop on any hierarchical level: For instance, in Figure 3, the meta-note around time 13 second in the improvisation is formed on an intermediary hierarchical level. So far, these subordinate events consist of notes forming the local group. But a richer understanding of the structural configuration is that the subordinate events of a local group consist actually of the smaller local groups that belong to that group. And this hierarchical structuration continues recursively for the smaller local groups. Thus when searching for the subordinate events that have same pitch than the last note of the group, we don't need to remember all the notes in the group, but only to check the pitch of the heads of the local groups one level down in the hierarchy.

3.2. Passing Note

Within a local group, all notes do not have the same importance. A monotonous and uniform conjunct melodic motion is a series of notes such that:

1. inter-pitch intervals between success notes are all of 1 or 2 semi-tones and in the same direction (up or down),
2. inter-onset intervals between successive notes are very similar,
3. no note is particularly accentuated.

In such configuration, the intermediary notes form *passing notes*: these subordinate elements play mainly a role of filling the interval gap between the starting and ending points of the line. For that reason, these intermediary notes are generally not perceived as note that play a more global role outside that particular melodic line. Intermediary notes are shown in grey in the graphical representations in Figures 3 and 4. For instance, in the first local group in Figure 3, in the ascending movement (C,D,E), note D plays a simple role of passing note between notes C and E. The longer descending movement around time 12 seconds (D,C,B,A,G) features a series of passing notes. Note "passingness" characterization has an impact in different modules of the integrated analysis framework. For instance, in the previous paragraph dealing with local group's head (section 3.1), a note within a local group that has same pitch than the last note of the group is excluded from the group's head if it is a passing note.

3.3. Syntagmatic Network

Melodic "reduction" is commonly understood as a process of eliminating the ornamentation (here, the subordinate elements) from the music surface in order to keep the deeper structure (on various hierarchical levels) made of the more important notes. Our conception of melodic reduction, however, does not impose such reduction of information, but on the contrary, integrates the deeper structure

[7] Temporal distance between successive notes corresponds to the rhythmic value of the first note, defined with respect to the second note. When facing an actual performance without actual score representation, the notes are temporally located at time onsets defined on a continuous axis in seconds. In this case, the temporal distance between successive notes corresponds to the Inter-Onset Interval (IOI).

information within the music surface. More precisely, the surface is formalized as a chain of connections between successive notes, i.e., a chain of *syntagmatic connections*, or a *syntagmatic chain*, following Saussure (1916)'s terminology. The deeper structure can be represented by adding new syntagmatic connections between successive elements in the deeper hierarchical levels. We obtain hence a *syntagmatic network* presenting a set of possible alternative syntagmatic chains. Formalized heuristics rule this construction of syntagmatic chains, as shown below.

The head of any local group is syntagmatically connected to the most recent note preceding the group that is not a passing note, as well as to the head of any local group closed by that note. For instance, in Figure 4, the first 3 notes are C#, D and C#. The second and third notes form a local group, with head C#. This head is then syntagmatically connected to the previous note, i.e., the first C#. Similarly, the head of any local group is syntagmatically connected to the note succeeding the group. In a series of passing note, there is a direct syntagmatic connection between the notes just before and after that series. For instance in Figure 3, in the longer descending movement around time 12 seconds (D,C,B,A,G), there is a direct connection between the start (D) and ending (G) notes of this line. A syntagmatic chain of notes of same pitch form a single meta-note whose time onset is set at the first note. This meta-note can be syntagmatically connected to any note following the meta-note. For instance, in Figure 4, we mentioned the syntagmatic connection between the first and third note. Having same pitch, these two notes form a single meta-note. This formalizes the cambiata around pitch C# materialized by the first 3 notes. This concept of syntagmatic network follows Lartillot & Ayari (2012), but we propose here a much simpler network grounded by stronger perceptual heuristics, based on local grouping and ornamentation reduction.

4. MOTIVIC ANALYSIS

As previously mentioned – and as will be further discussed in the next sections – *parallelism*, i.e., the detection of sequential repetitions, plays a core role in music analysis. It plays not only a central role in motivic/thematic analysis, but can also be used to model metrical analysis in particular. Yet 20th century musicology has shown that parallelism is one of the mechanism that is the most difficult to formalize. Lerdahl and Jackendoff (1983) emphasize the core importance of parallelism, and acknowledge that they cannot integrate such complex mechanism in their theoretical framework. Computational research too struggles to establish models that could automate the detection of sequential patterns.

We have developed our own approach (Lartillot, 2005) that offers the following breakthroughs: Multidimensional sequential pattern repetitions can be adaptively detected from the whole set of musical dimensions: a pattern is a repetition of a sequence of descriptions, where each successive description can be expressed on various musical dimensions. For instance, in Figure 4, the whole excerpt is a repetition of 2 phrases, which are indicated by the upper red line at the bottom of the figure: the left graduation on the line indicates the beginning of the first phrase, the middle graduation separates the first and second phrases, and the right graduation terminates the second phrase. This first half of the repeated pattern is a very specific melodico-rhythmical sequence (C#,D,C,#,E,E; C,D#,C,D,D), whereas the second half is a more general rhythmical sequence formed by a succession of the rhythmical cell (fourth note, eighth note) repeated three times.

The pattern analysis can be performed "exhaustively", without loss of information, but in the same time the results can be displayed in a compact representation. This allows an "exhaustive" representation of all the motives discovered. "Exhaustive" here means that there is no global filtering of the results that would discard information that might turn out to be of importance. Still the model includes necessary mechanisms that prevent combinatorial explosions of structural inferences (for instance due to the successive repetition of a same pattern). These mechanisms are based on interesting heuristics (for instance based on cyclic representation of repeated patterns) that were discovered as part of this research, and that seem to relate to actual cognitive descriptions of the way human listeners process structure while listening to music.

The analysis is performed chronologically from the start to the end of the piece of music, note after note. This perception-like modeling enables to formalize the interactions between the pattern detection module and all the other modules that can exist at any time while listening to the piece, note after note. For instance, the ornamentation reduction analysis described in previous section progressively builds a syntagmatic network on the piece of music under analysis. The new syntagmatic connections are fed to the pattern discovery module as they correspond to possible expansions of the sequential patterns. In this way, sequential pattern repetition can be detected in polyphony through voice streaming, as well as in the presence of ornamentation or reduction. The newly discovered pattern extensions can be further used for grouping and streaming. For instance, in Figure 4, the second half of the aforementioned phrase (upper red line at the bottom of the figure) is also characterized by a melodic sequence (A,B,C#,D,C#,B), which is completed by different types of ornamentation on each occurrence of the phrase, but that can be tracked on one specific syntagmatic chain within the syntagmatic network[8]. Motivic patterns are not composed of a mere sequence of note and interval descriptions: their internal structure is often characterized by the presence of subpatterns. Modelling this important aspect requires the formalisation of patterns of patterns, which is currently under study.

[8] We may notice here that this melodic sequence would require a syntagmatic connection between C# and D around time 6 second, which remains absent. This connection could be grounded by a new heuristic to be formalised, adding syntagmatic connections based on *pattern expectation*: when the pattern is extended up to note C# on both occurrences, the next note D in the second occurrence (around time 13 second) is considered as a possible syntagmatic continuation of the first occurrence (around time 6 second).

5. METRICAL ANALYSIS

Adequate rhythmical description of notes requires the inference of the underlying pulsation, and more generally of multiple pulsation levels forming a metrical structure. Current beat tracking methods are based on global description of periodicity (such as autocorrelation function) (Dixon, 2007), which, here also, fails to grasp particular idiosyncrasies of music, such as ornaments or sudden changes of tempo.

Periodicities can be expressed as successive repetitions of sequential patterns. The pattern discovery method described in the previous section is applied among others to detect pulsation and construct the metrical structure (Lartillot, 2010). The metrical structure cannot always be inferred from a mere search for periodicity in the signal, but may sometimes require the recognition of learned rhythmical patterns. This study is currently under investigation.

6. MODAL ANALYSIS

Previous methods in tonal and modal analysis traditionally compute global pitch statistics (of either whole pieces or on successive arbitrary time frames) that are compared to mode or key templates (Krumhansl, 1990; Gomez, 2006; Gedik & Bozkurt, 2010). The main limitations are that these arbitrary time frames often encompass complex modulations or spurious note events that are foreign to the main mode, and that the templates force a single stereotypical representation of each mode. In contrast, we are conceiving a new paradigm for modal analysis carried out for each successive note in the proto-symbolic representation, and based on a comparison of the local context (defined by the current and recent notes, and taking into account ornaments reduction) with all possible modes and key scales. A scale can also be identified while recognizing only subset of it, by taking also into account pivotal notes in the scale and longer and main notes in the local context. A first model of modal analysis dedicated to traditional Maqam music is detailed in (Lartillot and Ayari, 2014). Current research attempts to enrich the model and generalize to the study of tonal music as well.

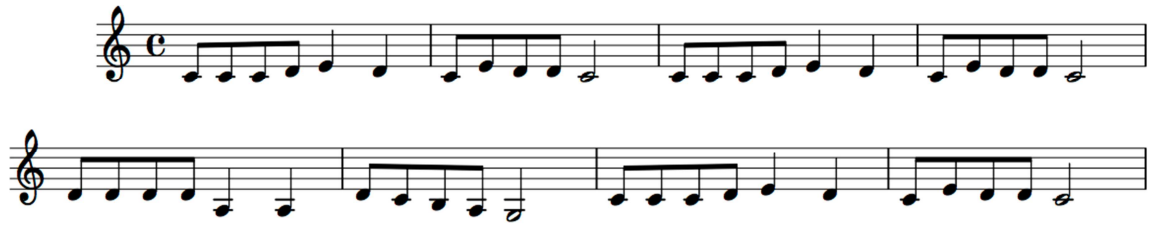


Figure 1: 18th-century French folk song "*Au clair de la lune*", analyzed in Figure 3.

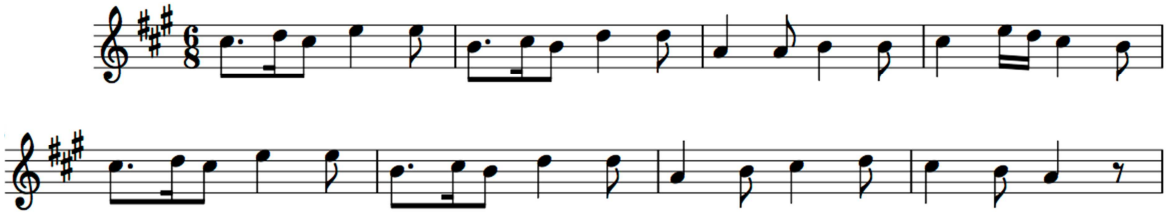


Figure 2: Beginning of the upper voice of the theme of the *Andante grazioso* of Wolfgang Amadeus Mozart's *Piano Sonata No. 11* in A major, K. 331, analyzed in Figure 4.

7. DISCUSSION

Figures 3 and 4 show the analysis produced by the computational framework, for two very simple musical examples shown in Figures 1 and 2. These simple examples have been chosen for the sake of simplicity, and due to the current limitations of the model under development. Still these simple analyses already show clearly the structure of these pieces. The French folk song (Figure 3) has a very clear structure with a tight congruency of the local grouping and the pattern repetition. On the contrary, the Mozart's *Sonata* theme (Figure 4) reveals an interesting contrast between the local grouping and the melodic and metrical structure. Analysis on traditional Tunisian improvisation in Maqam style is shown in (Lartillot and Ayari, 2014).

The computational framework under development is made freely available9, as a package, called *MusMinr*, of *MiningSuite*, a new *Matlab* environment for audio and music analysis (Lartillot, 2011). Besides the practical application to automated music analysis, the computational model, by virtue of its generality, extensiveness and operationality, can be offered as a blueprint for the establishment of cognitively validated model of music structure apprehension.

9 This Matlab toolbox can be downloaded from the website at the following address: http://code.google.com/p/miningsuite

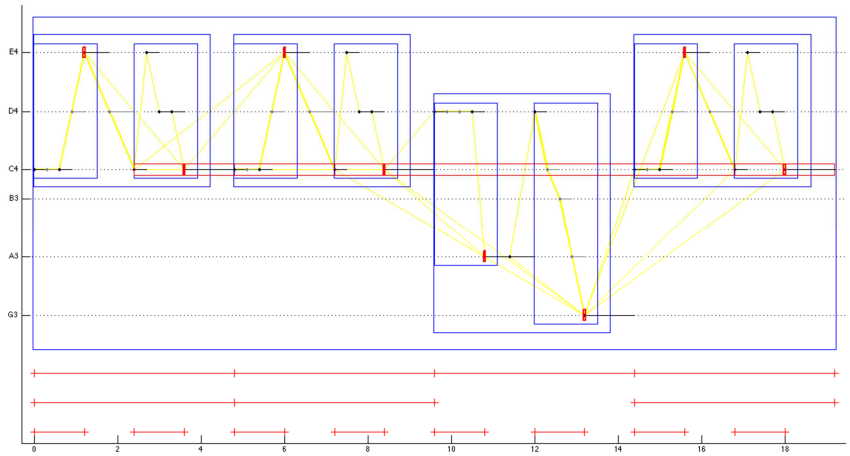


Figure 3: Analysis of the 18th-century French folk song "*Au clair de la lune*" shown in Figure 1. The excerpt is shown in piano-roll representation, with time in X-axis and pitches in Y-axis. Each note is indicated by a black (or grey) diamond followed by a black (or grey) line indicating its duration. Local groups are squared in blue. Local group heads are squared in red. Passing notes are shown in grey. The syntagmatic network is shown in yellow. The motivic pattern analysis is displayed below the piano-roll representation by red lines. Each row corresponds to a particular pattern and each occurrence is indicated by a line bounded by two graduations, which are aligned to the first and last note of that pattern occurrence in the piece of music above.

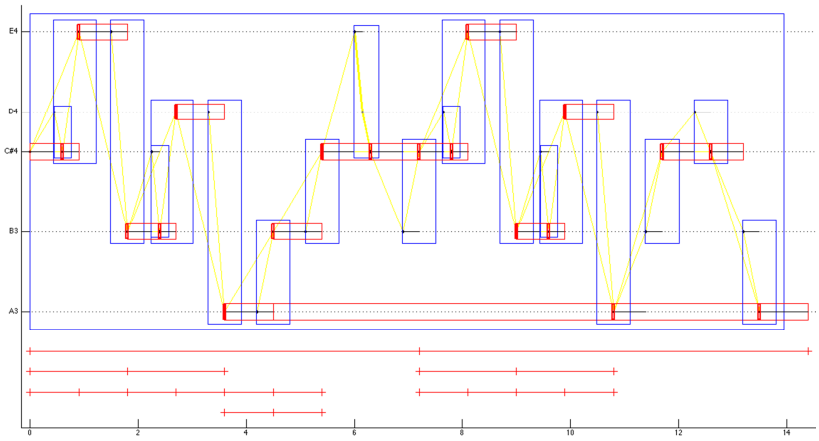


Figure 4: Analysis of the beginning of the upper voice of the theme of the *Andante grazioso* of Wolfgang Amadeus Mozart's *Piano Sonata No. 11* in A major, K. 331, shown in Figure 3.

8. ACKNOWLEDGMENTS

This research has been funded by Olivier Lartillot's Academy of Finland research fellowship, carried out at the Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä. This paper and the future research are funded by the European project *Learning to Create* (Lrn2Cre8), which acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859. This study has benefited from stimulating discussions with Mondher Ayari (in the context of our *CréMusCult* project, funded by the French research agency (ANR) during the years 2011- 2013), Mathieu Giraud, Petri Toiviainen and Funda Yazıcı.

9. REFERENCES

Cambouropoulos, E. (2006). Musical parallelism and melodic segmentation: A computational approach. Music Perception, 23*(3), 249–269.*

Dixon, S. (2007). Evaluation of the audio beat tracking system beatroot. Journal of New Music Research, 36*(1), 39–50.*

Gedik, A. C. & Bozkurt, B. (2010). Pitch frequency histogram based music information retrieval for turkish music. Signal Processing, 10*(1049-1063).*

Gilbert, E. & Conklin, D. (2007). A probabilistic context-free grammar for melodic reduction. In Proceedings of the In- ternational Workshop on Artificial Intelligence and Music, IJCAI-07.

Gomez, E. (2006). Tonal description of music audio signal. *PhD thesis, Universitat Pompeu Fabra, Barcelona.*

Krumhansl, C. (1990). Cognitive foundations of musical pitch. *Oxford University Press.*
Lartillot, O. (2005). Multi-dimensional motivic pattern extraction founded on adaptive redundancy filtering. Journal of New Music Research, 34*(4), 375–393.*
Lartillot, O. (2010). Reflexions towards a generative theory of musical parallelism. Musicae Scientiae, Discussion Forum 5, *195-229.*
Lartillot, O. (2011). A comprehensive and modular framework for audio content extraction, aimed at research, pedagogy and digital library management. In Proceedings of the AES 130th Convention.
Lartillot, O. & Ayari, M. (2012). Retentional syntagmatic net- work, and its use in motivic analysis of maqam improvi- sation. In Proceedings of the 2nd International Workshop of Folk Music Analysis.
Lartillot, O., & Ayari, M. (2014). A comprehensive computational model for music analysis, applied to maqam analysis. In Proceedings of the 4th International Workshop on Folk Music Analysis.
Lartillot, O., Yazıcı, F., & Mungan, E. (2013). A pattern- expectation, non-flattening accentuation model, empirically compared with segmentation models on traditional turkish music. In Proceedings of the 3rd International Workshop on Folk Music Analysis.
Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music. *Mass: M.I.T. Press.*
Marsden, A. (2010). Schenkerian analysis by computer: A proof of concept. Journal of New Music Research, 39(3), *269-289.*
Nattiez, J.-J. (1987). Music and Discourse: Toward a Semiology of Music *(Carolyn Abbate, Trans.) New-Jersey: Princeton University Press. (1979).*
Reti, R. (1951). The Thematic Process in Music. *Westport, CT: Greenwood Press.*
Saussure, F. de (1916). Cours de linguistique générale. *[Course in general linguistics], Paris: Payot.*
Tenney, J. & Polansky, L. (1980). Temporal gestalt perception in music. Journal of Music Theory, 24, *205–241.*

[3D] Social Psychology of Music

LIÈGE 08:30- 10:30 Tuesday 5 Aug 2014

IMPACT FACTORS ON THE MUSICAL IDENTITY OF MUSIC STUDENTS, PROFESSIONAL MUSICIANS, AND MUSIC TEACHERS

Wilfried Gruhn[1,3], Kristi Kiilu[1], Reet Ristmägi[1], Karin Täht[2]
[1]Estonian Academy of Music and Theatre, Tallinn, Estonia, Estonia, [2]University of Tartu, Estonia, Estonia, [3]University of Music Freiburg, Germany, Germany
Time: 08:30

ABSTRACT

Based on findings from former research on musical identity and self-concepts a cross-sectional study was designed on the development of individual differences which arise from education and the particular job requirements in music performance and education. The relevant research questions pertain to the identification of (1) main factors that govern identity formation, (2) developmental changes in the process of identity formation and (3) individual and group differences. Developmental changes over time might be supported by the activities and challenges of a particular occupation and the completed training programs. Therefore, music students from different programs (performance vs education at music academies vs universities) were investigated and compared with older, experienced professional musicians. All participants were measured by a Musical Identity Scale (MIS) which contains musical, educational, social and personal components. Additional information about cognitive, personal, and musical abilities was collected from a non-verbal cognitive test (Raven's Progressive Matrices), a standardized personality scale (Neo-5 Factor Inventory), and a music aptitude test (Gordon's Advanced Measures of Music Audiation, AMMA). The results from a multivariate analysis (MANOVA) and a factor analysis of the four dimensions of MIS exhibit a clear and significant differentiation between groups and their educational background. Four main factors could be extracted that influence identity formation.

1. INTRODUCTION

Musical identity is a psychological construct that consists of various components which are mainly based on personal, social, emotional and musical competencies, and influences the self-concept, self-esteem, and self-efficacy of individuals. MacDonald and collaborators differentiate between different "identities in music" (IIM) and the different function of "music in identities" (MII) (MacDonald, Hargreaves, et al., 2002). This points to an important and necessary distinction in terms of the function of music during the process of identity formation. Here, we only deal with identities developed in different musical and social contexts.

Former research has shown that musical identities reflect the individual understanding of the "closeness" of an activity to the self: how much does one feel that an activity shapes the self (Kessels & Hannover, 2004). Based on this theoretical approach Maria Spychiger and collaborators have developed a multidimensional scale for their empirical investigation of identity building factors (Spychiger, Gruber, et al., 2009). Their analysis exhibited that cognitive components perform the strongest impact on the musical identity in professional musicians and music workers while the more "spirituel" aspects were most prominent in amateurs (ebd., 3). Other studies have focused on teacher identities (Ballantyne, Kerchner, et al., 2012; Welch, Purves, et al., 2010) and on cross-cultural studies (Green, 2011a; b). The study presented here looks for the dynamic aspects of the development of identity formation during education and professional experience. It seeks to identify the impact factors that govern this process and the implications for education and teacher training programs. The main interest is to answer three questions:

1. What are the main factors that determine the identity formation in musicians?
2. Which processes can be observed through age and can be associated with developmental changes?
3. What are the specific traits within different groups according to their training and professional background?

2. METHOD

A total of N = 119 subjects participated in the study. Since development is a crucial issue of the study a cross-sectional design was performed. Young music students from different programs (performance vs education at music academies vs universities, N = 47, mean age 25,5 years) and older professional musicians with at least 10 years of professional experience (performing musicians vs teachers, N = 60, mean age 44,9 years) were compared with regard to their scores exhibited in all tests. This provided the opportunity to investigate differences within and between groups for each component of the MIS. Finally, a group of pupils (N = 12, mean age18,5 years) of a special music high school (Musikgymnasium) was added which served as a control group.

All participants filled in an online questionnaire indicating personal data (age, gender, education, status, socio-economic status of the family, family background). Furthermore, subjects were asked to fill in the Music Identity Scale (MIS) which was specifically designed for this study and modifies the basic idea of Spychiger's Scale of Musical Self-Concept (Spychiger et al., 2009). The MIS consists of 40 statements regarding personal (P),social (S), educational (E) and musical (M) behavior. The subjects ranked all statements according to a 6-point Likert scale. These components shape four-dimensional profiles of individuals, age groups and groups according to their occupation or training which, then, can be analyzed and compared. Additionally, measurements of cognitive advancement, musical aptitude and personality factors were taken from the last two parts of Raven's Standard Progressive Matrices (Raven, 1990), Gordon's Advanced Measures of Music Audiation (AMMA) (Gordon, 1989) and the short version S5 of Costa & McCrae's NEO Five Factor Personality Inventory (Costa & McCrae, 1992; Konstabel, Lönnquist, et al., 2012). MIS, the Raven Matrices and the Personality Inventory (S5) were presented online.

All data were statistically analyzed with SPSS 22. A multivariate analysis (MANOVA) and a factor analysis of the four dimensions of MIS were performed. For the group comparison a two-sample t-Test was employed.

3. RESULTS

The data present an ambivalent picture which reveals partly unexpected results while others conform to plausible expectations. Since both cohorts (students and professionals) constitute a rather homogeneous selection of musically active and highly trained subjects, no relevant differences can be found with regard to instrumental training and familiar social background. However, a comparison of the particular groups reveals enlightening findings. No significant differences exist between school music teachers and instrumental teachers with respect to music aptitude and cognitive scores, personality factors and identity components. However, music education and performance students who mark the situation about 20 years before professional job establishment exhibit significant differences in the cognitive Raven test ($p = .039$), for the personality factors extraversion ($p = .023$) and agreeableness ($p = .041$) and for the educational component of MIS ($p = .039$). This unfolds a surprising development from student to professional music teacher. While a person works as a teacher for several years his/her attitude becomes strikingly more educational whereas the period of their studies creates different life perspectives, professional visions and artistic expectancies. That indicates that professional involvement in the occupation of a musician as teacher or performer impacts the development of the actual musical identity.

This is also confirmed by the results of the comparison between performers (professional orchestra players, choir singers) and music teachers in public schools. These groups perform significant differences in music aptitude ($p = .001$), cognitive development ($p = .002$), extraversion ($p = .003$) as well as regarding the social ($p = .035$) and the educational ($p = .017$) components of MIS. The multivariate analysis of all tested dimensions exhibits a significant correlation only for musical ability scores and cognitive development ($r = .334$).

Therefore, the profiles of the different groups which are derived from the MIS data draw a clear picture of the dominant components regarding the development of characteristics for each subject group (see figure 1).

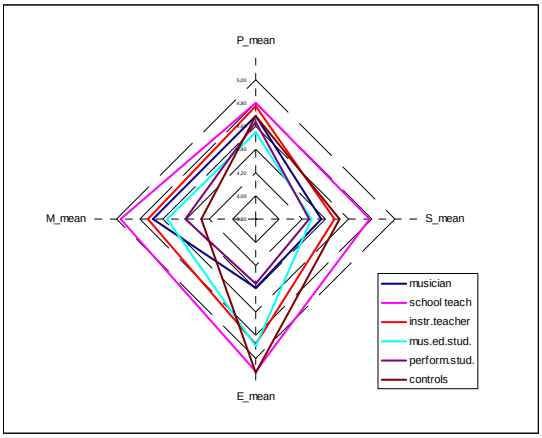


Figure 1: Profiles of the four components of MIS for all groups. P = personal; S = social; E = educational; M = musical.

Here it is obvious that music teachers in public schools exhibit the most pronounced profile with highest consent to all four components whereas professional musicians and performance students show less explicit profiles with lower values of agreement. Instrumental teachers position themselves between both groups and exhibit average means for all four components. The most obvious difference occurs for the controls which are still pupils without a clear social and musical profile.

Differences also occur with respect to the personality factors. Professional musicians and school teachers significantly differ regarding their development of the extraversion factor. The same results are demonstrated in the two student groups (performers and educators; see figure 2). It is also highly illuminating that the neuroticism factor shows the largest distribution and the most pronounced characteristics for all groups. The psychological disposition for an extrovert or neurotic behavior apparently refers to a discrete impact factor for musical identity as teacher or performer.

A factor analysis of the data for all components of MIS (personal, social, educational, musical) exhibits four main factors that determine the MIS profile. These factors can be interpreted as theoretical reflection (F1), social interaction (F2), personal relation (F3) and emotional affection (F4).

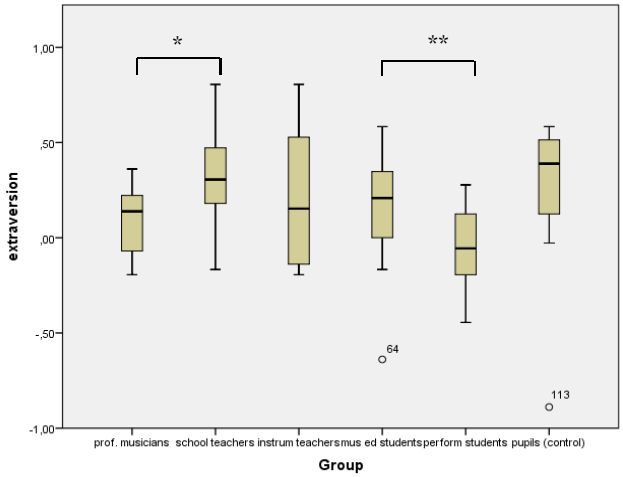


Figure 2: The extraversion dimension of the personality scale for all groups. Differences for professional musicians and teachers as well as for the two student groups are highly significant (**): musicians vs teachers p = .009; performance vs education students p = .017.

The factor analysis of the data of all groups demonstrates remarkable differences. Professional musicians strongly load on F4 (emotional affection) whereas school music teachers most prominently load on F2 and F3 (social interaction and personal relations). This does not at all surprise, it rather confirms what can be expected as a result of training and experience in combination with the respective personality factors e.g. (extraversion). However, only music education students most strongly load on F1 (theoretical reflection). (see figure 3).

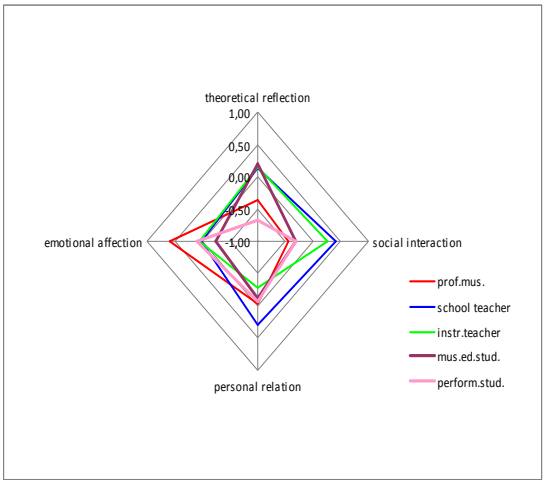


Figure 3: Loading factors of the factor analysis.

A comparison of the factors for professional musicians and school music teachers demonstrates the quite opposite characteristics of both groups (figure 4). All differences are highly significant.

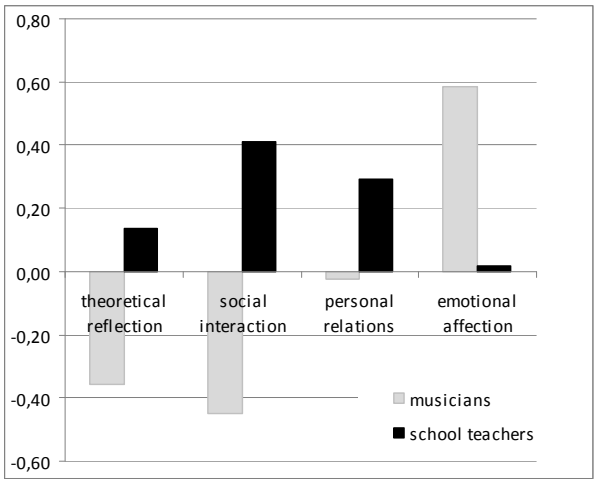


Figure 4: Loading factors for performing musicians and school music teachers.

4. DISCUSSION

The study has demonstrated that specific impact factors develop in different dimensions and differ between groups. It has also become apparent that personality factors as well as chronological dimensions of age and the amount of time spent with a professional activity (occupation) are highly relevant for the identity formation. A clear and pronounced attitude and self-concept occurs in subjects according to their training and their vocational devotion to be a performer or teacher. Here, it is especially notable that not only or predominantly the genetically determined personality traits are responsible for the development of performers or teachers, but rather the environmental and educational, personal and institutional context shapes the musical identity over the years. This also indicates that not only the age determines a musical identity, but rather the time actively spent in an occupation which generates the typical attitudes and behaviors. The longer one works as a skilled and passionate performer or teacher the more one develops those attitudes and personality factors that are relevant for successful work in a profession.

However, we do not know to which extent the different factors interact. One could assume that the typical mixture depends on the personal psychological disposition. On the other hand, it is well observed and generally known in students that the program and the very personal, intense and intimate relation to an instrumental teacher in music students have an enormous impact on the adoption of attitudes, norms and values.

To explore these factors in a more general context, it would be necessary to include more subjects from domains other than music to step deeper into an understanding of the hidden processes and impact factors that are relevant to the formation of musical identity. This will be extremely important to institutions of higher education and teacher training to model and implement those educational programs that help to support abilities and areas of interest that are relevant to model musical identities through music and within the domain of musical activities. With respect to this demand one needs to contrast musicians and music educators with other students and professionals to unfold common principles and psychological structures within the developmental process.

5. REFERENCES

Ballantyne, J., Kerchner, J.L. & Aróstegui, J.L. (2012). Developing music teacher identities: An international multi-site study. International Journal for Music Education, 30*(3), 211-226.*

Costa, P.T.J. & McCrae, R.R. (1992). NEO Five Factor Inventory. *Odessa, FL.*

Gordon, E.E. (1989). Advanced Measures of Music Audiation (AMMA). *Chicago: GIA Publ. Inc.*

Green, L. (Ed.). (2011a). Learning, teaching, and musical identity: voices across cultures. *Bloomington: Indiana University Press.*

Green, L. (2011b). Musical identities, learning, and education: some cross-cultural issues. In B. Clausen (Ed.), Vergleiche in der musikpädagogischen Forschung *(pp. 11-34). Essen: Blaue Eule.*

Kessels, U. & Hannover, B. (2004). Empfundene Selbstnähe als Mediator zwischen Fähigkeitsselbstkonzept und Leistungskurswahlintentionen. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36*(3), 130-138.*

Konstabel, K., Lönnquist, J.-E., Walkowitz, G., Konstabel, K. & Verkasalo, M. (2012). The 'Short 5' (S5): Measuring personality traits using comprehensive single items. European Journal of Personality, 26, *13-29.*

MacDonald, R., Hargreaves, D.J. & Miell, D.E. (2002). Musical Identities. *Oxford: Oxford University Press.*

Raven, J.C. (1990). Standard Progressive Matrices. *Oxford.*

Spychiger, M., Gruber, L. & Olbertz, F. (2009). Musical Self-Concept. Presentation of a multi-dimensional model and its empirical analyses. In J. Louhivuori & T.E.S. Saarikallio (Eds.), Proceedings of the 7th Triennial Escom Conference *(Vol. 2009, pp. 503-506). Jyväskyla.*

Welch, G., Purves, R., Hargreaves, D.J. & Marshall, N. (2010). Reflections on the teacher identities in music education. Action, Criticism & Theory of Music Education, 9*(2), 11-32.*

MUSIC "VIOLENCE": HURT BEHIND THE SOUNDS

Zhuo Zhang[1]
[1]*Zhangzhuo, China*
Time: 09:00

On this world, we can hear music whenever and anywhere. Why we must to receive it? Sometimes, the music doesn't means graceful and fair-sounding and which can make us dysphoric and anxiety. To describe this, we call it Music Violence. This paper try to define music violence, and list its various forms. What is the music violence? It is the undesirable phenomenon in music, especially in pop music, who ignores the respect for the rights of people, makes use of the strengths of money and media, and last takes advantage of these Measures forcing people to receive voice of music and peoples can't refuse or avoid it. Definition of music violence include two forms: violence in music and violence of music. In some songs, it's lyric contain words about violence, for example, sex, war, murdur, drug and so on and in some Music Video,we can see blood, force, weapon. All of these, we call it violence of music.This paper main content is discuss about violence of music,include for part:

1. Backgroud: This part summarize the study about music violence before and give the concept of music violence.
2. Pychology effect of music violence: This part would to describe how music violence influence our life and how the sounds hurt us.
3. Music violence: theoretical explain and significant this part try to analysis phenomenon of music with social pychology, media psychology and cognitive pychology. Be a kind of art, music affect our emotion, even change our life through it's sound. In modern times, musical sounds combine with mass media and musical industy to build a huge space what makes people can't distinguish art or hurt. This paper would like to uncover the power and elusive of music violence.

INTERACTIONISM AND MUSIC: MEASURING ENTRAINMENT AND INFLUENCE IN DYADS AND GROUPS

Tommi Himberg[1], *Neta Spiro*[2]
[1]*Brain Research Unit, O. V. Lounasmaa Laboratory, Aalto University, Finland,* [2]*Research Department, Nordoff Robbins, United Kingdom*
Time: 09:30

Embodied and social cognition of music has drawn increasing interest and explorations of new approaches. These approaches redefine the cognitive system to extend beyond individual brains/minds. Although there is growing consensus that this redefinition is necessary for the understanding of complex, social behaviours such as music and dance, in practice it has turned out to be difficult. We suggest that one of the main obstacles has been the use of "traditional" methods, especially in measurements and statistical analysis. These methods tend to treat the phenomena as static and focus on the individual. They are thus inappropriate for the study of joint actions such as music-making in groups, in which the participants and the environment are in dynamic interaction. Though recent dyadic and group studies on music and dance include new methods, these are mostly exploratory and we are far from converging on methodological solutions and paradigms. Based on our review of such empirical and theoretical studies, we propose that adopting an interactionist (e.g. De Jaegher & Di Paolo, 2007) framework would help solve many of the current methodological problems. In this paper, we aim to 1) review the recent dyadic and group studies on music, focusing on methods, 2) present an interactionist framework for studying musical interaction, and 3) suggest methodological solutions to overcome problems in dyadic and group music studies. According to our review, the dyadic or group studies on music undertaken recently used several methods, often not optimal for social settings, leading to highly constrained or artificial tasks. In discussing the interactionist approach to social cognition in the context of music we suggest, for example, that directing the focus onto the dynamics of interaction rather than onto individual players would help resolve some of the issues of experimental control. In addition to a general framework, we provide practical suggestions for conducting dyadic and group studies. Social interaction is a key component of music-making in many contexts, and developing methods for its study in more natural settings will have implications for studies of music therapy, music education, and basic research on music psychology.

ALL THE LOVERS ARE TENORS: EVOLUTIONARY PSYCHOLOGY AND QUEER IN VERDI

Hasan Gürkan Tekman[1]
[1]*Uludag University, Turkey*
Time: 10:00

ABSTRACT

An examination of the distributions of voice types in the operas of Verdi, as a representative of nineteenth century opera, indicates two patterns: (1) The tenor wins the affections of the female lead and (2) Verdi heroines fall in love with a character sung by a tenor while in a long term relationship with a character sung by a baritone or bass. These patterns are in conflict with the findings of evolutionary psychology. Women are expected to be attracted to men who show evidence of high levels of testosterone and dominant social position.

These characteristics are found to be signaled by voices with lower fundamental frequencies and formant frequency patterns that indicate larger physical size, respectively. Furthermore, the female preference for deeper male voices is found to be mainly related to short-term mate selection. Thus, we should expect that attractive males should be vocally characterized by deeper voices and women in a permanent relationship with tenors should seek short-term liaisons with baritones. While research in evolutionary psychology fails to explain the representation of the desirable male in opera, we may turn to arguments regarding how opera, and music in general, represents gender and sexuality. First, music can be characterized by a gendered discourse that maintains a patriarchal order on the one hand and finds ways of subverting the gender categories of this order on the other. Second, operatic tradition in particular has been indifferent to established gender categories as evidenced by both male and female roles written for both male (castrato) and female singers with female voices throughout the seventeenth and eighteenth centuries. Third, both men and women of unconventional sexualities have been attracted to opera as composers, performers, and fans from the nineteenth century onwards. The prevalence of the tenor voice in the male lead during the nineteenth century, which is taken as the most heteronormative period of this genre, can be interpreted as an extension of an attitude to gender and sexuality that we may call queer today.

1. BACKGROUND AND AIM

The main aim of this paper is to discuss the contrast between the representation of the male hero in the operas of Verdi and the kind of male voice that is found to be attractive to women in research done within the framework of evolutionary psychology. Considering this is the kind of opera that continues to be most popular in live and recorded performances of opera, this contrast between the accepted vocal portrayal of the desirable male and the findings of evolutionary psychologists must be relevant to how gender and sexuality is interpreted by present day opera listeners. Then, approaches that emphasize the socially constructed nature of gender are discussed as an alternative to the evolutionary psychology approach to the presentation of the male hero in opera. In order to frame these questions in greater detail a short discussion of the casting of different voice types in male and female roles in the history of opera appears necessary.

1.1. Voice Types in Opera

Transition from castrato/travesti to tenor. When opera emerged as an art form at the end of the 16th century assigning the role of the main protagonist to a castrato (a male singer castrated before puberty in order to prevent him from going through change of voice) was quickly established as a convention. Two reasons are usually offered for this convention (Abbate & Parker, 2012; Evans, 2005). One is a difference in the perception of gender compared to the current Western idea that gender is a dichotomous classification: Until the nineteenth century the gender was considered as a continuum from male to female rather than two discreet classes in Europe. The second reason is the fact that typically protagonists in operas were monarchs or princes who went through a crisis that usually involved a conflict of duty and love. The royal authority invested by God to such a figure made him above human categories such as gender: Such a character could be portrayed on stage by a singer who had the vocal range of a woman and the physical appearance that combined masculine and feminine features.

As the fascination with castrati, which enabled some castrati to acquire immense fame and fortune during the 17th and 18th centuries, wore off and they started to be seen as a kind of freak, the voice type of the main male protagonist had to change at the beginning of the 19th century. For a few decades vocal range of the leading male was still a woman's (André, 2006). However, a woman dressed as a man (*en travesti*) would take on this role. This was a custom that already had a long history because even at the time the opera stage was dominated by castrati, a woman *en travesti* would take on a male part if a castrato was not available. By the 1830s however, tenors claimed the part of the male protagonist on the operatic stage and that remained the tradition to this day. Apart from the diminishing acceptance of female voices in male roles at this time, the changes in singing technique that resulted in a powerful higher extension being added to the tenor voice was a main cause of this change.

Portrayal of the feminine in opera. This transformation in how the male protagonist was portrayed in opera influenced the status of female characters as well. André (2006) argues that the period from approximately 1800 to 1830 when some women singers sang both female parts and male parts *en travesti*, contributed to the development of female characters in opera during the rest of the 19th century. The pattern, seen clearly in many Rossini and early Bellini and Donizetti operas before 1930, casts a female lead as a woman who faces trials and tribulations because of her love for the male lead and usually dies for it. This is contrasted with a second part for a woman's voice, which could be either a male lead sung by a woman or a second woman who is usually a rival and does not die at the end of the opera. Later in the 19th century (later Bellini and Donizetti and Verdi are good examples) most operas have one rather than two leading parts for women. These single leads combine aspects of the two parts for women's voices in a single opera that was seen earlier.

1.2. Question of the Tenor Hero

The portrayal of men on the operatic stage by women's voices has been a topic of discussion in literature on opera (Bashant, 1995; Dame, 2006). This issue has even been related to the construction of femininity in opera after the period of male protagonists with female voices have come to an end (André, 2006). Tenors as the natural representation of the male hero in opera, on the other hand, seem to have been accepted without much discussion. However, findings of evolutionary psychologists go contrary to casting high male voice as the target of the female protagonist's affections.

Promising mates according to evolutionary psychology. According to evolutionary approaches to mate selection people should be genetically programmed to be attracted to individuals who will maximize the chances of survival and reproduction of the offspring they may have together (Buss, 2012). Regardless of gender, individuals with high quality genes (indicated by symmetrical appearance, freedom from parasites and illness, etc.) should be preferred as potential males. Command (or the potential to have command) over resources and the willingness to invest in a long-term relationship with a specific mate and in the offspring produced within that

relationship are additional qualities that women look for when selecting a mate. The position of a man in the social dominance hierarchy is related to the resources that he may possess and pass on to his offspring. Potential to achieve positions of dominance is related to certain physical traits that are displayed visually: A large and muscular body with a strong upper body has indicated strong hunting prowess and ability to dominate other males in the environment of selection.

However, a large and strong body is also signaled by the kind of voice that that body produces (Evans, Neave, & Wakelin, 2006). Men with larger bodies produce voices with lower formant frequencies and less separation between the formant frequencies. Formants are frequency bands that contain higher amplitudes of vibration and are determined by the size of the cavities that are formed in the vocal tract by the placement of the chin/ tongue and the lips. Furthermore, lower fundamental frequency of a voice, which is related to the size of the vocal folds, indicates a greater influence of testosterone. Thus, a deeper voice with lower fundamental frequency and formant frequencies signal presence of qualities associated with greater dominance. An evolutionary psychology hypothesis would be that women should be attracted to men with these vocal qualities.

Feinberg, Jones, Little, Burt, and Perrett (2005) verified the prediction about a female preference for men whose voices had lower fundamental frequencies. Feinberg and others electronically manipulated the acoustic characteristics of the voices of young males rather than using natural voices with low or high fundamental frequencies. This was because men with naturally high or low fundamental frequencies could have other acoustic or speaking characteristics that could have been confounded with fundamental frequency. Feinberg, DeBruine, Jones, and Little (2005) also found that women's preferences for vocal masculinity were correlated with their preferences for facial masculinity.

Long-term and short-term mating. Another aspect of the evolutionary psychology approach to women's mate selection is related to the context of long-term and short-term mate selection. Mastery of resources and the commitment that a man makes to a woman and her children are more important in selecting a long-term mate because those characteristics would directly influence the chances of the survival and reproductive success of the children. However, in short-term mate selection the more important qualities become those that make a man physically attractive. If the father is not likely to contribute to caring for the offspring/ the reproductive success of the male offspring depends on inheriting physical characteristics that women find especially attractive: Known as the sexy-son hypothesis (Gangestad & Simpson, 1990). To support this idea, it was found that estrogen levels were associated with women's preferences for masculinized male voices. Whereas women in the fertile phase of their cycle of ovulation showed greater preference for masculinized voices than women in other phases (Feinberg et al., 2006; Puts, 2005), women using oral contraceptives showed this effect to a smaller extent (Feinberg, DeBruine, Jones, & Little, 2008). Furthermore, the effect of phase in the ovulatory cycle was greater in the short-term mate selection context (Puts, 2005).

2. THE TENOR HERO IN NINETEENTH CENTURY OPERA AND VERDI

Any person even remotely familiar with opera would be aware of the fascination with the tenor voices and the typical casting of the tenor as the ardent lover. This tradition that is in direct contradiction with the kind of male voice that women should find attractive will first be examined in greater detail in the works of one of the most popular opera composers. Then, an alternative approach to opera and gender will be taken up.

2.1. Verdi

The tenor voice as the impersonation of the operatic hero is mainly a tradition of the 19th century. Guiseppe Verdi is one of the best known and most frequently performed opera composed who lived and worked during the 19th century. In fact, he is one of the best known and most frequently performed opera composer of any time (Abbate & Parker, 2012).

Verdi's career straddled a turning point in the history of opera (de Van, 1992): He started off when opera composers were considered artisans who delivered popular entertainment to audiences that demanded a continuous flow of new works. Impresarios brought such works to production and they had the main control over the profits of the business. Singers also had considerable control over productions and, as a result, over composers because of their fame and ability to bring in audiences. During the later parts of Verdi's career, however, opera had started to be viewed as an art form characterized by works that would make up a canonized repertory that would receive repeated productions. Composers became artists who had the potential to become known for posterity. This increased the amount of control that composers had on productions of their works. Furthermore, publishers emerged as a force in opera because they owned a considerable share in the rights of the works.

Although he saw such drastic changes in the way operas were produced during his career, the way Verdi treated different voice types shows consistency throughout his career. This stable pattern was representative of how voices were distributed to the characters in Italian opera from 1830s to early 20th century. The stability of this casting tradition across the transformation that opera went through in the 19th century shows that it was not related to the status of opera as popular entertainment or high art, but had to stem from other sources.

2.2. The Tenor Voice in Verdi Operas

According to de Van (1992) the "hero" is the character type most closely associated with the tenor voice in Verdi. The hero loves and is loved in return by the soprano. He is usually in conflict with figures of authority and he obeys to a single law to the end, that is, the law of love.

Tenors as objects of affection in Verdi. If we actually count the male protagonists who earn the affections of a female character who appears in the opera in the 28 Verdi operas, we find 30 characters cast as tenors and only 4 cast as baritones or basses. Thus, we are

able to see the magnitude of the preference for men with tenor voices on the part of the female characters of Verdi operas. This definitely does not fit the expectation derived from the evolutionary psychology research.

Another way in which we may examine the reproductive success of operatic characters is by looking at the probability that they actually produce offspring. If we consider the male characters of Verdi who we know to have children, we find a considerable advantage of deeper voices in this category. We find 24 baritones or basses in comparison with only 3 tenors who appear as fathers in the 28 operas. Thus, in this respect the distribution of voices in Verdi operas seems to fit the evolutionary psychology hypothesis that men with deeper voices should have better success in fathering children. However, deeper voices are also usually associated with greater age in Verdi. These baritone and bass fathers are usually fathers of adult children, whereas most tenor characters are young men who may become fathers in the future - if they survive the end of the opera, that is. If we count the male protagonists in Verdi operas who remain alive at the end of the drama and has the potential to become fathers, we find 14 tenors and 5 baritones or basses. Considering they have most of their reproductive years ahead of them, Verdian tenors do not fare badly in terms of the potential to produce offspring either.

To sum up, men who attract women as romantic partners and who are likely to have children and transmit their genes to the next generations are more likely to be portrayed by tenors than by baritones and basses in the operas of Verdi. The informal observation that "all the lovers are tenors" bears out in closer inspection.

Tenors as lovers of married women in Verdi. According to the viewpoint of evolutionary psychology, commitment to a spouse or a relationship is one quality that women look for in selecting mates. Thus, women may have to settle for men who have less promising genes or lower positions in the dominance hierarchy because that is balanced by their willingness for a committed relationship. Considering that voice characteristics primarily carry information about testosterone level, size, and dominance they may not be strongly predictive of long-term mating preferences. This was verified by evolutionary psychological research as well (Puts, 2005). One reason Verdi and other composers of romantic opera cast the tenor as the desirable romantic partner may be they thought of their heroines as choosing long-term mates. In that case operatic women may not be necessarily attracted to deep voiced men. Thus, short-term mate selection may be a better place to look for a correspondence between evolutionary psychology and operatic voices.

Short flings with men on the part of a heroine is not a typical plot device in Verdi, although *Rigoletto* and *La Traviata* offer exceptions to this. However, some Verdi operas include love triangles in which a married woman is attracted to another man. This appears as a promising case to compare long-term and short-term mating preferences of operatic women because we can consider the voice types of the husband and the extramarital flame.

Five Verdi operas (*La Battaglia di Legnano*, *Stiffelio* and its later revision *Aroldo*, *Un Ballo in Maschera*, and *Don Carlos*) contain such triangles. Except for *Stifellio* and *Aroldo*, in which both men are tenors, the husband is a baritone or bass (in *Don Carlos*) and the lover is a tenor. Significantly, only in the common plot of *Stifellio* and *Aroldo* the heroine is in love with her husband and the extramarital affair is the result of a trick played on her. In all the other operas the soprano falls (or was) in love with the tenor but remains faithful to the baritone or bass husband for sake of duty. Thus, tenors are not the reliable albeit not so attractive men that women settle for in order to secure the livelihood of their children, but rather the dashing men who sweep them off their feet. This again is in contrast with the predictions of evolutionary psychology.

3. A QUEER PERSPECTIVE

Representation of the male protagonist in opera does not conform to the kind of male voice that is found desirable by women in evolutionary psychology research as can be seen in examination of the 28 operas by Giuseppe Verdi as a typical case, above. Neither the pattern of attractions to males for long-term and short-term mating consistent with the viewpoint of evolutionary psychology. Of course factors such as the acoustic and musical capabilities of higher voices and the possibility of two lovers' voices intertwining and coming together on the same note, which is drastically reduced as the pitch ranges of the two voices become more separate, may have contributed to the preference for higher voices being cast as the romantic partners of women in opera. Furthermore, it may be argued that opera is not an art form that aims for realism. First, the fact that the characters sing rather than speak requires a radical suspension of disbelief from the opera audience. Second, and more closely related to the subject of this paper, the visual representation of operatic characters are not very realistic either. On the one hand, most of the time middle-aged singers with hefty physique portray teenagers in love. On the other, the acting of opera singers can miss realism because of the acting styles of the period or the limitations of singers as actors. Nevertheless, I will go on to consider the possible contribution of a queer perspective to this issue.

Queer theory is a viewpoint that emphasizes the socially constructed nature of gender and sexuality in order to challenge established categories and norms. Judith Butler (1990), one of the central thinkers within queer theory, emphasizes that even the concept of biological sex is a dichotomy that is constructed and substituting "gender" in place of "sex" does not help if the assumption is that every "male" will turn into a "man" and every "female" will turn into a "woman". Butler views sex, gender, and desire as open to alternative constructions that do not align with these dichotomies. Hence, a queer perspective promises to relate to the representation of gender and sexuality in opera, which does not comfortably fit into biologically based expectations about gender and sexuality.

Gender and sexuality in music. From the time of Plato and Aristoteles music has been considered subversive in Western thought in its ineffability and resistance to being brought into a rational frame (Brett, 2006). From the time science and medicine started to define homosexuality, its possible connection with music and art has been discussed (Brett & Wood, 2006). McClary (1990) argues that bringing this irrational aspect of music within the control of a patriarchal order has been one constant strand in the history of Western music. The gendered classifications (the "masculine versus feminine" analogies for dualities such as major/minor, tonal/atonal, diatonic/chromatic, etc.) that are common in music theory anchored music in masculinity and used femininity to describe instability.

McClary (1990) believes that the "feminine" sides of such dichotomies have been used in music in order to depict departures from the masculine and the rational. This may include homosexuality (McClary, 2006) as well as femininity (McClary, 1990).

Gender and sexuality in opera. Unlike symphonic music or chamber music, opera is a genre of Western music that has no claim to be an abstract source of esthetic pleasure. Fictional persons appear on stage physically and make their voices heard in an opera. Hutchinson and Hutchinson (1996) have addressed questions that relate directly to bodies such as disease, death and desire in the realm of opera. Furthermore, authors such as Brett (2006) and Evans (2005) have found parallels between homosexuality and music and opera. Just as homosexuals are marginalized in society, music is marginalized as an art form and opera is marginalized in music.

With its tradition of castrati at its very beginnings, opera can be said to have a subversive attitude to gender. Apart from (or maybe in combination with) acoustic reasons for having castrati sing the roles of the hero, the penetrating and overpowering quality of the castrato voice may have contributed to its ability to portray a male in the vocal range of the female (Dame, 2006). This created the possibility of a character that was both male and female in addition to the more obvious neither male, nor female interpretation (Bashant, 1995) interpretation. The blurring of gender boundaries was further enhanced by the possibility of castrati singing male or female roles. Dame (2006) adds one more complication to the subversive nature of opera in term of gender by pointing out that the penetrating "masculine" voice of castrati put the listeners of either sex in the feminine position by the feelings of awe and surrender it inspired.

Although the time of the castrati came to an end early in the 19th century on the opera stage, operas appeal to unconventional sexualities continued in other forms. Both lesbians and gay men found a special attraction to opera in Western countries with strong traditions of opera. This is probably most visibly demonstrated as the type known as the "opera queen" (Evans, 2005; Koestenbaum, 1993). However, several scholars have discussed the connection between opera and lesbians as well.

Blackmer and Smith (1995) recount the anecdote of an anonymous lesbian who felt herself attracted to snippets of opera that she heard, first by chance, then by choice, while she was an undergraduate in the 1970s in USA. This woman was amazed at her later discovery that women could dress up as men in order to sing male roles who often were lovers of female characters. The rather confusing pleasure she felt in the conjunction of two female voices in duets led to the surprising discovery of a realm, which she imagined as belonging to upper-class, reactionary people, where women in drag could sing love duets with other women and make a show of it.

Wood (2006) coined the term "Sapphonics" in order to describe both the aspect of the opera experience that she connected with lesbian desire and the voices of women who sing (sang) characters that are either men *en travesti* or women with such strength and determination that is considered typical of men. Wood cites examples such as the early 19th century singer Pauline Viardot-Garcia, late 19th century singers Emma Calvé and Olive Fremstad, and early 20th century singer Mary Garden. Bashant (1995) adds another late 19th century singer, Johanna Wagner to this list. These singers crossed the boundaries between soprano and alto voice types by force of will, to sing repertories that were characterized by roles that were either men or women who violated the social limits set for women at the time. Viardot-Garcia and Fremstad inspired novels that featured characters modeled after them by George Eliot and Willa Cather, respectively. These works included themes of sexual attraction by a woman directed at a woman that was stimulated by the singing voice.

Another striking example of Sapphonics is the turn of the century lesbian composer Ethel Smyth (Wood, 1995; 2006). Although not a professional singer, Smyth played her operas on the piano and sang the parts with her voice that crossed boundaries of the voice categories in order to promote her works by conductors and producers. Wood (2006) points to our attention both the unconventional female characters that populate Smyth's operas and her juxtaposition of two female voices or one male and one female voice in similar registers in duets that suggest same-sex attraction.

"Opera queen" is a popular term in North American gay jargon that refers to American gay men who have an intense interest that reaches the point of an obsession in opera. Koestenbaum (1993) drew many parallels between gay men, opera, and divas in a book entirely about opera queens. One characteristic of opera queens is that they have very strong attachments to particular female opera singers to the exclusion of others. They express both their adoration of their favorite diva and their denigration of her rivals in florid and irreverent language (Morris, 1993). Koestenbaum sees being a diva as similar to being gay in the sense that both first become aware of this side of their identities as an internal feeling which is then followed by personal acceptance and public acknowledgement. Another aspect of the diva (and opera in general) that resembles at least a segment of gay men is the degree of excess in its expression. This word, "excess" is one of the most popular to describe many things in opera from the elaborate stage sets of the 17th and 18th centuries to intricate coloratura ornamentation used to portray insanity to eccentricity taken to uncomfortable levels such as in Salome's long address to the severed head of John the Baptist in Strauss' *Salome* (Abbate & Parker, 2012). This has a parallel in the dress, mannerisms, and language that typifies certain gay subcultures (Evans, 2005).

Koestenbaum (1993) goes on to find more symbolic relationships between opera and gay culture that has to do with the appearance of the opera queen early in the 20th century. At his time new works that joined the standard repertory became a rarity and recorded operas became available. Koestenbaum finds a parallel between the confinement that appears in listening to recorded opera (doing it at home, the singer's voice being captured within the grooves of a vinyl disk, images of singers being fit within the frame of a black disk in the advertisements of Columbia records in Opera News magazine) and being closeted. Furthermore, he sees the vinyl record as a mirror that offers the chance for looking for a secret self in the voice that in contains. Finally, opera as a dead art in addition to the almost compulsory death of the protagonists at the end from early 19th century onwards may have a special appeal for gay audiences. Taylor (2002) points out that portrayals of homosexuals as self-destructive and doomed to an early death has become a stereotype that gay men have internalized themselves.

4. CONCLUSIONS AND IMPLICATIONS

Representing gender and desire can be argued to be central to music. Throughout history music has been used to subvert gender and sexual categories. By its subversive nature representing sexual attraction in a realistic manner may not be a priority of opera. Thus, the vocal casting of the desirable male hero as the tenor rather than the baritone or bass, which should have been more attractive to women, makes sense although it is inconsistent with predictions of evolutionary psychology.

There are three points to which I would like to draw attention related to this conclusion. First, this relationship between music and more specifically opera and queer may be interpreted in different ways. Although Brett (2006) sees this relationship as kind of a bribe given by society to a subset of gay men in exchange for not making themselves visible in unacceptable ways, Evans (2005) thinks stereotypes such as the opera queen may be used in ways to challenge socially constructed views of homosexuality. Even the position that opera, music, the arts in general are associated with gay sensibilities can be given essentialist (that the psychological makeup of gay men underlie artistic creativity) or constructionist and queer interpretations.

Second, this conclusion rests upon information that has been obtained through both strictly empirical research, such as the characterization of the effect of voice qualities on women's mating preferences, and theoretical work that takes a more critical standpoint as its foundation. Although these two approaches are usually seen as alternatives that exclude each other, much may be gained from their conjunction in a single work.

Finally, this approach brings up the question whether gay and lesbian viewpoints may have more to say in music psychology. Gay and lesbian sensibilities have found their voice in both psychology and musicology in the early 1990s. The collections that have been brought together by Blackmer and Smith (1995b) and Brett, Wood, and Thomas (2006) bear witness to the contribution that considering the sexualities of composers, performers, and listeners can make to the understanding of music. Although perception and cognition of music has been a subject that saw considerable developments in importance during the same period, homosexuality, and gender in general, has received negligible attention. If some people feel that their sexual orientation is an important part of their experience of music, music psychology could benefit from taking gender and sexuality in consideration in understanding why and how people engage with music.

5. REFERENCES

Abbate, C. & Parker, R. (2012). A history of opera: The last 400 years. London: Allen Lane.

André, N. (2006). Voicing gender: Castrati, travesty, and the second woman in early-nineteenth-century Italian opera. *Bloomington: Indiana University Press.*

Bashant, W. (1995). Singing in Greek drag: Gluck, Berlioz, George Eliot. In C. E. Blackmer & P. J. Smith (Eds.) En travesti: Women, gender, subversion, opera. *New York: Columbia University Press.*

Blackmer, C. E. & Smith, P. J. (1995a). Introduction. In C. E. Blackmer & P. J. Smith (Eds.) En travesti: Women, gender, subversion, opera. *New York: Columbia University Press.*

Blackmer, C. E. & Smith, P. J. (1995b) En travesti: Women, gender, subversion, opera. *New York: Columbia University Press.*

Brett, P. (2006). Musicality, essentialism, and the closet. In P. Brett, E. Wood, and G. C. Thomas (Eds.) Queering the pitch: The new gay and lesbian psychology, *2nd ed. New York: Routledge.*

Brett, P. & Wood, E. (2006). Lesbian and gay music. In P. Brett, E. Wood, and G. C. Thomas (Eds.) Queering the pitch: The new gay and lesbian psychology, *2nd ed. New York: Routledge.*

Brett, P., Wood, E., and Thomas, G. C. (2006) Queering the pitch: The new gay and lesbian psychology, *2nd ed. New York: Routledge.*

Buss, D. M. (2012). Evolutionary psychology: The new science of the mind. *Boston: Allyn & Bacon.*

Butler, J. (1990). Gender trouble. *New York: Routledge.*

De Van, G. (1992). Verdi: Un théatre en musique *[Verdi: A theater in music]. Paris: Fayard.*

Dame, J. (2006). Unveiled voices: Sexual difference and the castrato. In P. Brett, E. Wood, and G. C. Thomas (Eds.) Queering the pitch: The new gay and lesbian psychology, *2nd ed. New York: Routledge. In P. Brett, E. Wood, and G. C. Thomas (Eds.)* Queering the pitch: The new gay and lesbian psychology, *2nd ed. New York: Routledge.*

Evans, D. T. (2005). Speaking over and above the plot: Aural fixation, scopophilia, opera, and the gay sensibility. Theory, Culture, & Society, 22, *99-119.*

Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between voice characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. Biological Psychology, 72, *160-163.*

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Little, A. C. 2008). Correlated preferences for men's facial and vocal masculinity. Evolution and Human Behavior, 29, *233-241.*

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E. Hillier, S. G., & Perrett, D. I. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. Hormones and Behavior, 49, *219-222.*

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. Animal Behavior, 69, *5761-568.*

Gangestad, s. W. & Simpson, J. A. (1990). Toward an evolutionary history of females' psychosexual variation. Journal of Personality, 58, *69-96.*

Hutchinson, L. & Hutchinson, M. (1996). Opera: Desire, disease, death. *Lincoln: University of Nebraska Press.*

Koestenbaum, W. (1993). Queen's throat: Opera, homosexuality, and the mystery of desire. *Boston: Da Capo Press.*

McClarey, S. (1990). Feminine endings: Music, gender, and sexuality. *Minneapolis: University of Minnesota Press.*

McClarey, S. (2006). Constructions of subjectivity in Schubert's music. In P. Brett, E. Wood, and G. C. Thomas (Eds.) Queering the pitch: The new gay and lesbian psychology, *2nd ed. New York: Routledge.*

Morris, M. (1993). Reading as an opera queen. In R. A. Solie (Ed.) Musicology and difference: Gender and sexuality in music scholarship. *Berkeley: University of California Press.*

Puts, D. A. (2005). Mating context and menstrual phase affects women's preferences for male voice pitch. Evolution and Human Behavior, 26, *388-397.*

Taylor, G. (2002). Psychopathology and the social and historical construction of gay male identities. In A. Coyle & C. Kitzinger (Eds.) Lesbian & Gay Psychology: New perspectives. *Oxford: Blackwell Publishers.*

Wood, E. (1995). The lesbian in the opera: Desire unmasked in Smyth's Fantasio *and* Fête Galante. *In C. E. Blackmer & P. J. Smith (Eds.)* En travesti: Women, gender, subversion, opera. *New York: Columbia University Press.*

Wood, E. (2006). Sapphonics. In P. Brett, E. Wood, and G. C. Thomas (Eds.) Queering the pitch: The new gay and lesbian psychology, *2nd ed. New York: Routledge.*

[3E] Power of Music

LOS ANGELES 08:30- 10:30 Tuesday 5 Aug 2014

EFFECTIVENESS OF SOUND EFFECTS AND MUSIC TO INDUCE LAUGH IN COMICAL ENTERTAINMENT TELEVISION SHOW

Ki-Hong Kim[1], Shin-Ichiro Iwamiya[2], Fumiya Mori[3]

[1]Faculty of Media and Information Resources, Surugadai University, Japan, [2]Faculty of Design, Kyushu University, Japan, [3]Tis Inc., Japan

Time: 08:30

ABSTRACT

The aim of the present study was to clarify the effect of sound effects and music on the inducing laugh of funny scenes of comical entertainment television shows. When the affective impression of an imagery sound, one kind of sound effect, was funny, the sound made the affective impression of video funnier when the sound was combined. On the contrary, a boring imagery sound made the affective impression of video more boring. An exaggeration sound, the other kind of sound effect, made the impression of video funnier when the video and sound was congruent regardless of funniness of the sound itself. In the case of symbolic music, when the meaning associated with music and content expressed in the video was matched, the affective impression of video became funnier.

1. INTRODUCTION

In television programs, various sound effects and music are combined to funny scenes to induce more laugh of viewer-listeners. Alten (2010) remarked that sounds play an important role for emphasizing the ridiculousness of a character in a scene to create humor. The purpose of the present study was to clarify the effect of sounds which were combined with videos on the inducing laugh of the videos.

To explore the effect of sounds used in funny scenes, Mori, Kim, and Iwamiya (2011) conducted a rating experiment using sounds and videos in Japanese television programs. They found that background music (BGM) and sound effects reinforced funny impression of video scenes. Furthermore, Mori, Kim, and Iwamiya (2012) suggested that funniness of *imagery sound* and BGM directly emphasized the funny impression of video while *exaggeration sound* and *symbolic music* increased funniness of video by the combination of a sound and visual element. To empirically prove these assumptions, we conducted a systematic experimental study using imagery sounds, exaggeration sounds, and symbolic music and examined how those sounds induced laugh of funny scene in video.

2. EXPERIMENT1: MEASUREMENT OF AFFECTIVE IMPRESSION OF AUDITORY STIMULI

2.1. Method

Firstly, we performed a series of rating experiments to show the features of sound effects and music, which would be used in the second experiment.

Experimental stimuli were 12 sound effects and eight music. Sound effects are pure tones (S4, S6), a glide FM tone (S2) whose carrier wave was sinusoidal and its frequency was linearly rose on a log frequency axis with periodically modulated by sinusoidal wave, glide pure tones in which frequency linearly rose on a log frequency axis (S3:100 to 500 Hz, S5:500 to 2,000 Hz), and the sound effects (S1, S7, S8, S9, S10, S11, S12) which were excerpted from a commercially available sound effect library (the sound effects series 10, CD KICG-175) and the sound effects available on the website (http://sonicwire.com/fxbgm). Among these sound effects, S1~S8 were used as imagery sounds and S9~S12 were used as exaggeration sounds. In the present study, "imagery sound" means the sound which cannot be made by the content of video; there is not actual connection between the sound source and the visual object. On the other hand, "exaggeration sound" is the sound which is supposed to be emitted from the visual object, however features of the sounds are exaggerated; there is connection between the sound source and the visual object. The musical stimuli are shown in Table 1. These stimuli were used as symbolic music, which might induce association of specific meanings. The symbolic music means that they have been continuously used for the specific purpose and we recall the associative meaning when we hear the music.

Table 1: Associative meaning (multiple answers were allowed) and listening experience rate (multiple choices were not allowed) for each music stimulus

Music No.	(Title)	Listening Experience rate (%)	Association Meaning (frequency)
M1	(Eye of the Tiger)	90	battle (9)，fighting sport (5)，man (4)，boxing (2)，entrance (2)，variety show (2)，fisticuffs (1)，match (1)，training (1)
M2	(Danger Zone)	75	battle (4)，fighter aircraft (4)，car (4)，television commercial (4)，rock music (4)，airplane (2)
M3	(FIFA Anthem)	75	soccer (14)，entrance (6)，FIFA World Cup (4)，athletic meet (4)，progress (4)，Japanese national team (3)，live telecast (2)
M4	(Waka Waka)	5	dance (5)，festival (3)，night club (3)，live performance (3)，ethnic (3)，aborigine (2)，southern country (1)，Africa (1)，South America (1)，soccer (1)
M5	(When You Wish Upon A Star)	100	night sky (10)，starlit sky (10)，Disney (8)，lullaby (7)，music box (6)，child (6)，sleep (4)
M6	(J. Brahms's Lullaby)	40	child (9)，lullaby (8)，sleep (7)，music box (3)
M7	(Pump It)	90	man (3)，black person (3)，variety show (3)，TAXI (the title of the movie) (2)，racing (2)，dash (2)，sense of speed (1)
M8	(Truth)	95	Formula 1 (11)，racing (4)，Ferrari (2)，car (2)，rivalry (3)，circuit (2)，exhaust sound (2)

Table 2: Excerpts of video clips as the visual stimuli and their features

Video No.	Excerpt (Title)	Content of Video Clip
V1	Japanese TV program (WORLD GREAT TV)	Dad with his baby has finished sliding on water slider in a pool, and then they were hit by a stranger who has slipped next.
V2	Japanese TV program (WORLD GREAT TV)	An old man was enjoying golf at a golf course, but he missed to drive a golf ball and fell down on the ground.
V3	YouTube (Standing & Walking Toy Poodle Part 1)	A toy poodle was walking with two legs by an upright posture such as a kangaroo on road of a residential street.
V4	YouTube (Midafternoon Walk of Chi Tiger)	A character of a tiger costume was walking along the mountain path pleasantly.
V5	Japanese TV program (Man! You're Screwed GP)	A retired boxer tried to dodge successively flying ping-pong balls but he could not, because the speed of the ball was too fast. He was directly hit by all of them.
V6	Japanese TV program (WORLD GREAT TV)	A young man was suddenly hit by a balance ball thrown by his brother in the stairs and the passage of their house.
V7	World CM Festival 2010 (Seriously Strong Cheddar)	A mouse ate cheese of the set on a mouse trap and then it was caught, but it started bench press using an iron bar of the trap.
V8	Japanese TV program (WORLD GREAT TV)	A soccer goal keeper suddenly appeared at a bowling alley. He started to block a bowling ball thrown from a person.
V9	Japanese TV program (WORLD GREAT TV)	A young girl fell asleep on a chamber pot halfway through a toilet.
V10	Japanese TV program (WORLD GREAT TV)	A young boy rode in a sled. His pet dog pulled it. They resplendently slid up and down on the snow-covered hill.

The rating experiment was performed in a soundproof room under the standard lighting condition. All the sound stimuli were stereophonically presented to the subjects via an audio interface (CardDeluxe CDX-01) of PC (Dell OptiPlex 745) and headphones (STAX Lambda Nova Classic System). The experimental task of the subjects was to rate the affective impressions of the sound stimuli using the eight kinds of seven steps, bipolar, adjective scales. All the sound stimuli were presented to the subjects in random order. In the case of musical stimuli, in addition to the rating experiment, the subjects were asked to answer the questions such as “have you ever heard the music?” and “what do you associate with the music.” In the survey regarding whether the subjects have heard the music or not, a forced-choice method was used. The subjects selected one of the three categories; [I have heard it], [I might have heard it once], and [I have never heard it]. For the survey on the association caused by music listening, a free-description answer was required.

The subjects were 20 Japanese undergraduate and graduate school students (15 males and 5 females) in the age ranged from 20 to 26 years old. All the subjects had normal hearing and vision (including corrected vision).

2.2. Results

The obtained data were applied to factor analysis using principal factor method. The scales were the variable. Three factor solution was obtained. They were interpreted as the *funniness factor*, the *impact factor*, and the *safety factor* depending on the factor loadings for each rating scale. Furthermore, the average factor scores of each sound stimulus on each factor were obtained.

According to the funniness factor scores of each stimulus, the sound effect stimuli were ordered, from boring to funny impression: S4 (the most boring), S6, S10, S12, S9, S3, S11, S8, S5, S2, S1, and S7 (the funniest). The music stimuli were also ordered, from boring to funny impression: M6 (the most boring), M5, M1, M2, M8, M3, M7, and M4 (the funniest). In addition, the listening experience rate on each musical stimulus (the percentage of 'I have heard it before'), the description on association with each music stimulus and its frequency were given in Table 1. Table 1 showed that most subjects have heard the music excepts M4 and M6. Table 1 also showed that the music stimuli M1 and M2 were associated with "battle," M3 with "soccer," M4 with "dance and festival," M5 with "night sky and starlit sky," M6 with "child and lullaby," M7 with "man and black person," and M8 with "Formula 1 and racing." These results confirmed the music stimuli in Table 1 aroused common association among subjects, and these music could function as symbols of associated meanings.

3. EXPERIMENT 2: EFFECTIVENESS OF SOUNDS ON THE FUNNINESS OF VIDEO CLIPS

3.1. Method

Next, we performed rating experiments to clarify the effectiveness of sound effects and music to induce laugh in videos. As shown in Table 2, ten kinds of funny videos as visual stimuli were selected from the television programs broadcasted in Japan, the DVD of the world CM festival 2010 (Best selection Vol. 1), and popular video websites (YouTube). We removed all sound components, such as dialogues, music, and sound effects, of the extracted audiovisual materials. The sound stimuli combining to the video stimuli were the same as used in the first experiment. The audiovisual stimuli were made by combining video clips and sound effects or music as the following. The video clips V1 and V2 were combined with the sound effects S1, S2, S3, and S4. The video clips V3 and V4 were combined with the sound effects S5, S6, S7, and S8. The video clips V5 and V6 were combined with the sound effects S9, S10, S11, and S12. The video clip V7 was combined with music M1 and M2. The video clip V8 was combined with music M3 and M4. The video clip V9 was combined with music M5 and M6. The video clip V10 was combined with music M7 and M8. The contents of videos V7 to V10 were supposed to match with the associations of combined music.

The rating experiments were performed using 10 kinds of visual stimuli (i.e., video clip alone) and 32 kinds of audiovisual stimuli (i.e., combination of video clip and sound effects or music). The experimental task of the subjects was to rate the affective impression of each visual stimulus and audiovisual stimulus using the rating scales of nine adjective pairs shown in Table 3. In addition, the perceived congruence between sound and moving picture was measured in the case of the audiovisual stimuli.

Twenty subjects who had participated in the first experiment also participated in the second experiment. The experimental devices, environment were the same as the first experiment.

Table 3: Factor loadings for each rating scale after varimax diagonal rotation.

Rating scale	Factors	
	Funniness	Safety
boring - funny	.88	.30
humorous - humorless	.82	.27
bad - good	.78	.34
powerless - powerful	.75	-.11
unpleasant - pleasant	.68	.48
sad - happy	.58	.51
serious - silly	.53	.33
common - unique	.53	.05
dangerous - safe	.02	.78
Contribution (%)	44.2	17.4

3.2. Results

Similar to the first experiment, the obtained data were applied to factor analysis. The scales were the variable. Two factor solution was obtained. Table 3 represented the principal factor loadings for each rating scale. The two factors were interpreted as the *funniness factor* and the *safety factor*. Furthermore, the average funniness and safety factor scores for each visual stimulus and audiovisual stimulus was obtained. The average factor scores for the stimuli were represented in the two-dimensional space in Figure 1. The orthogonal axes of the Figure 1 represented the funniness and safetyfactor scores. The difference of the positions between visual-only stimuli and audiovisual stimuli using the same video clips indicate the effectiveness of sound effects and music on the affective impression of the video clip. If the funniness factor scores increase by combining sounds, these sounds are effective to induce laugh for videos.

In the case of the imagery sounds, as shown in Figure 1 (a) and (b), the affective impression of audiovisual stimuli became funnier when the sounds were combined with the videos than when only the video clips were presented in the cases of V1+S1, V2+S1, V3+S5, V3+S7, V4+S5, and V4+S7. On the contrary, the affective impression of audiovisual stimuli became more boring when the sounds were combined than when only video clips were presented in the cases of V1+S4, V2+S4, V3+S6, and V4+S6. According to the results of the first experiment, the affective impression of the sound effects S1, S5, and S7 was very funny, while that of S4, S6 was very boring. The affective impression of imagery sounds might directly influenced that of video clips. Funny sounds made the videos funnier, and boring sounds made the videos more boring. A linear relationship was supposed to be between the funniness factor scores of imagery sounds and the changes of the funniness scores from visual-only condition to audiovisual condition of each video clip. Actually, a strong positive correlation between them ($r(14)=.83$, $p<.01$) was observed.

In the case of exaggeration sounds, as shown in Figure 1 (c), similar to the case of imagery sounds, the funny sound S11 made the video clips funnier. This trend was shown by the changes of funniness factor scores from V5 and V6 to V5+S11 and V6+S11, and S11 was the only one exaggeration sound whose affective impression was funny. However, in the case of exaggeration sounds, even when the affective impression of the sounds was boring, the affective impression of video clips became funnier when the sounds were combined. This trend was shown when V5, V6 were combined with S9, S10, respectively. In the case of exaggeration sounds, rather than the direct effect of funniness of sounds, the connection or congruence between sounds and video clips might be more important to increase funniness. Actually, the affective impression of V6+S12 was more boring than that of V6, and the congruence score between the sound and video clip was the lowest ($M=1.7$, $SD=1.2$). The correlation coefficient between the average congruence scores of each video clip and the changes of the funniness scores from visual-only condition to audiovisual condition of each video clip was statistically significant ($r(6)=.83$, $p<.05$). Correlation coefficient between the funniness factor scores of exaggeration sounds and the changes of the funniness factor scores from visual-only condition to audiovisual condition of each video clip was not statistically significant ($r(6)=.43$, $p>.01$). Exaggeration sounds were effective to increase funniness when the sounds were recognized appropriate for the content of video clips.

In the case of symbolic music, as shown in Figure 1(d), the affective impression of all the video clips was much funnier when the music was combined with the videos than when only the video clips were presented. The results of first experiment showed, the affective impression of all the musical stimuli was boring except for M4 and M7. In the case of symbolic music, similar to the case of exaggeration sounds, the direct effect of funniness of sounds might not be effective. The connection between music and video clips was supposed to be more important to increase funniness. The associative meaning aroused from the music basically matched the content expressed in the scenes of video clips. In the case of symbolic music, the effectiveness of sounds depended on the relationship between associative meaning of music and the content of video clip. For instance, V8 was a video that expressed the ridiculousness of a soccer goal keeper appearing at a bowling alley. When M3 (FIFA Anthem) associated with "soccer" was combined with V8, the funny impression was improved. As another instance, V9 was a video that a young girl fell asleep on a chamber pot. When M6 (J. Brahms's Lullaby, Op. 49, No. 4) associated with "child and lullaby" was combined with V9, the funny impression was improved. Correlation coefficient between the funniness factor scores of symbolic music and the changes of the funniness factor scores from visual-only condition to audiovisual condition of each video clip was not statistically significant ($r(6)=.31$, $p>.01$). The funniness of music itself does not influence the funny impression of video clips. Namely, the matching between the associative meaning of music and the content of video clips contributed to enhancing the laugh in videos.

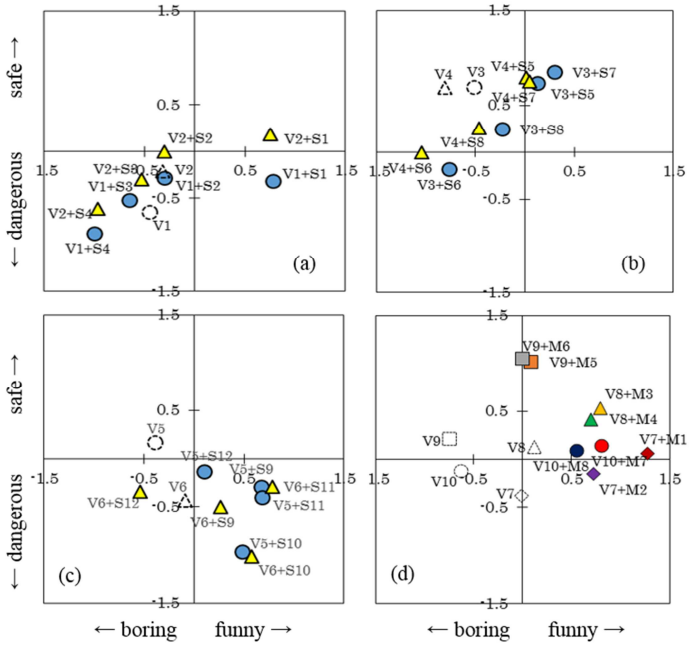


Figure 1: Factor scores for each visual stimulus alone (V) and their combinations with each sound stimulus (S: Sound effects, M: Music)

4. CONCLUSIONS

The effectiveness of sound effects and music that induce laugh in comical entertainment television show was investigated. Rating experiments clarified that sounds contributed to creating hummer of funny scenes of videos. The exaggeration sounds well matched funny scenes and the funny imagery sounds were effective to improve funniness of video clips. The symbolic music well induced associative meaning was also effective.

5. ACKNOWLEDGMENTS

This work was supported by Grant-in-Aid for Young Scientists (B) No. 25750002 & Grant-in-Aid for Challenging Exploratory Research No. 25560006 from MEXT Japan.

6. REFERENCES

Alten, S. R. (2010). Audio in Media. *Cengage Learning.*

Mori, F, Kim. K. H., & Iwamiya, S. (2011). Influence of Sound Effects and Background Music on Funniness of Comical Scenes in Television Programs. Proceedings of the Technical Committee on Musical Acoustics of the Acoustical Society of Japan, 74, *107-127. (in Japanese)*

Mori, F, Kim. K. H., & Iwamiya, S. (2012). Effect of Sound on Laugh Scene in TV Programs. Proceedings of the auditory research meeting, 42*(9), 739-744. (in Japanese)*

MUSIC AND ITS INDUCTIVE POWER: A PSYCHOBIOLOGICAL AND EVOLUTIONARY APPROACH TO MUSICAL MEANING

Mark Reybrouck[1]
[1]*Ku Leuven - University of Leuven, Belgium*
Time: 09:00

ABSTRACT

The aim of this contribution is to broaden the concept of musical meaning from an abstract and emotionally neutral cognitive representation to an emotion-integrating description that is related to the evolutionary approach to music. Starting from the dispositional machinery for dealing with music as a temporal and sounding art, musical emotions are considered as adaptive responses to be aroused in human beings as the product of neural structures that are specialized for their processing. A theoretical and empirical background is provided in order to bring together the findings of music and emotion studies and those of evolutionary musicology. The theoretical grounding elaborates on the transition from referential to affective semantics, the distinction between expression and induction of emotions, and the tension between discrete-digital and analog-continuous processing of the sounds. The empirical background provides evidence from several findings such as the infant-directed speech, the referential emotive vocalizations and separation calls in lower mammals, the distinction between the acoustic and vehicle mode of sound perception, and the bodily and physiological reactions to the sounds.

1. INTRODUCTION

Music is a powerful tool for induction of emotion and mood modulation by triggering ancient evolutionary systems in the human body. The study of emotions, however, has been introduced only recently as a topic in music research (Juslin & Sloboda, 2010; Trainor & Schmidt, 2003: Juslin & Västfjäll, 2008). Emotions, in fact, are difficult to study adequately and the same holds true for the idiosyncrasies of individual sense-making while listening to music. The development of new research methods (continuous, real time and direct recording of physiological correlates of emotion) and advanced techniques for medical imaging, however, have changed dramatically the field, with seminal contributions from the domain of neuropsychology, neurobiology, psychobiology and affective neuroscience. The domain has been broadened even further by encompassing contributions from evolutionary musicology, which claim an adaptive function for musical emotions.

2. MUSIC AS EXPERIENCE

Music can be considered at a structural or experiential level. The *structural level* conceives of music as a symbolic stimulus with semantic meaning that is dealt with “outside of time”. It is characterized by symbolic operations on mental replicas of the sounds, relying basically on processes of memory and imagination. As such, it brings about the concept of virtual simultaneity, which calls forth the transformation of a perceptual flux to a kind of object and the conception of much-at-once, as an extremely compressed, i.e., instantaneous or synoptic kind of representation (Godøy, 1997). We should thus conceive of a dynamics of representation that defines temporality as a modality of the music, which is not limited by the actual unfolding of the primary perception. Each of the velocities in our representations can provide a different kind of perspective and knowledge of the musical substance. Time, therefore, must be accepted as a morphological dimension of the musical object, somewhat analogous to the distinction Kramer (1988) has drawn between the *linear* or *active* and the *non-linear* or *still-spectator* mode of listening. The critical element in this distinction is the contiguity of the discrete particulars and their perceptual bonding. It is possible, in fact, to deal with the sonorous unfolding at the concrete level of sounding articulation, with elements which are presented to the senses, but it is possible also to deal with mental replicas of the sounds at a symbolic level of representation.

The *experiential level* of dealing with music (Reybrouck 2005) is dependent on perceptual bonding and continuous processing of the sounds. It involves a real time listening experience, which is both sensory-driven and time-consuming and is related to the body and its physiological responses as a 'theatre for emotional processing', somewhat reminiscent of James' controversial view on the mechanisms of feelings and emotions (Damasio, 1994, p.129) which, for short, can be stated as the postulation of a mechanism in which particular stimuli in the environment excite—by means of an innate set of inflexible mechanisms—a specific pattern of body reactions. In this view, the body is the main stage for enactment of emotions, either directly or via its representation in somatosensory structures in the brain.

3. MUSIC AND EMOTIONS

The domain of music and emotions has received recently much impetus from theoretical contributions and empirical research. There is, however, still need of a conceptual and theoretical framework that brings all findings together in a coherent way.

There is, first, a distinction between the recognition and experience of emotions. The former is a cognitive-discrete process, reducible to a categorical description of the sounds; the latter involves an experiential level of dealing with music, wich implies continuous rather than discrete processing of the sounds (Schubert, 2001; Nagel et al., 2007) and which is related to the distinction between the acoustic and vehicle mode of sound perception (Frayer and Nicolay, 2000). The "acoustic mode" involves emotive meaning with particular sound patterns being able to convey emotional meanings. They refer to the immediate, on-line emotive aspect of sound perception and production and deal with the emotive interpretation of musical sound patterns throughout "sentic modulation" (Clynes, 1977), which is a general modulatory system that is involved in conveying and perceiving the intensity of emotive expression by means of three graded spectra: tempo modulation, amplitude modulation, and register selection, somewhat analogous to the well-known rules of prosody. It refers to the local risings and fallings, quickenings and slowings, and loudenings and softenings that are involved in expressively communicating meanings. The "vehicle mode", on the contrary, involves referential meaning, somewhat analogous to the lexicosemantic dimension of language with particular sound patterns as vehicles to convey symbolic meaning. It refers to the off-line, referential form of sound perception and production, which is a representational mode of dealing with music that results from the influence of human linguistic capacity on music cognition (Frayer & Nicolay, 2000).

It can be argued, further, that the affective impact of music is generated by the modulation of sound with a close connection between primitive emotional dynamics and the essential dynamics of music. They appear to be biologically grounded and can be considered as innate release mechanisms that generate instinctual emotional actions (Burkhart, 2005; Hauser, 1996). Along with the evolved appreciation of temporal progressions (Clynes & Walker, 1986) they can generate, relive, and communicate emotion intensity, helping to explain why every basic human emotion is easily rendered through music.

The primary meaning of music, therefore, cannot be understood in terms of an abstract and emotionally neutral cognitive representation, but is dependent on emotion-integrating regions in the human brain (Brown, Martinez & Parsons, 2004; Menon & Levitin, 2005; Panksepp, 2009-2010). Musical semantics, therefore, is in search not only of the lexicosemantic but also of the experiential dimension of meaning, wich, in turn, is related to the affective one. Affective semantics—as the term has been coined—is able to recognize the emotional meanings which particular sound patterns are trying to convey. It calls forth an analog-continuous rather than a discrete-digital processing of the sounds in order to catch the expressive qualities that vary and change in a dynamic way. Emotional expressions, in fact, are not homogeneous over time, and many of music's most expressive qualities relate to structural changes over time, somewhat analogous to the concept of prosodic contours as they are found in vocal expressions (Scherer, 2003).

The strongest arguments for these claims come from the developmental perspective (Trainor & Schmidt, 2003): caregivers around the world sing to infants in an "infant-directed" singing style—using both lullaby and playsong style—which is probably used in order to express emotional information and to regulate their infant's state. This style (motherese) is distinct from other types of singing and young infants are very responsive to it. Additional empirical grounding comes from a typical class of primate vocalizations, which are coined as referential emotive vocalizations (Frayer and Nicolay, 2000) and separation calls (Newman, 2007). Embracing a body of calls that serve a direct emotive response to some object in the environment, they exhibit a dual acoustic nature in having both a referential and emotive meaning. As such, they are exemplary of the above-mentioned acoustic and vehicle mode of sound perception.

4. INDUCTION OF EMOTIONS: PSYCHOBIOLOGICAL CLAIMS

Music can be considered as something which catches us and which induces several reactions that are beyond conscious control, or as something that can be processed in a conscious way. There is, as yet, a deeper affective domain to which cognition is subservient and which makes the brains such receptive vessels for the emotional power of music (Panksepp & Bernatzky, 2002). Functional neuroimaging studies on music and emotion have shown that music can modulate activity in brain structures that are known to be crucially involved in emotion, such as the amygdala, the nucleus accumbens, the hypothalamus, the hippocampus, the insula, the cingulate cortex and the orbitofrontal cortex (Koelsch, 2014). The prime motivations for engaging with music, moreover, are experiencing and regulating of emotions and moods (Menon & Levitin, 2005).

Emotional reactions to music, further, activate the same cortical, subcortical and autonomic circuits, which are considered to be the essential survival circuits of biological organisms in general (Trainor and Schmidt, 2003). The subcortical processing affects the body through the basic mechanisms of chemical release in the blood and the spread of neural activation. The latter, especially, invites people to react bodily to music with a whole bunch of autonomic reactions—changes in heart rate, respiration rate, blood flow, skin conductance, brain activation patterns, and hormone release (oxytocin, testosterone)—, which are driven by the phylogenetically older parts of the nervous system (Ellis and Thayer, 2010). They can be considered to be the "physiological correlates" of listening to music and demonstrate convincingly the role of autonomic changes which are associated with emotion processing.

Music thus has the potential to evoke emotions, or put in other terms: it has *inductive power.* It can be subsumed under the effects, which are triggered by the central nervous system and which may engender physiological responses. The latter are proportional to the way the information has been received, analyzed and interpreted through instinctive, emotional pathways that are ultimately concerned with maintaining an internal environment that ensures survival for both the individual and the species (Schneck & Berger, 2010). This dynamically equilibrated and delicately balanced milieu (*homeostasis*), together with the physiological processes which maintain it, relies on finely tuned control mechanisms that keep the body operating as closely as possible to predetermined baseline physiological quantities or *reference set-points* (blood pressure, pulse rate, breathing rate, body temperature, blood sugar level, pH, fluid balance, etc.). Sensory stimulation of all kinds can change and disturb this equilibrium and invite the organism to adapt its basic reference points, mostly after persisting and continuous disturbances that act as environmental or driving forces to wich the organism must adapt. There are, however, also short term and immediate reactions to the music as a driving force, as evidenced from neurobiological and psychobiological studies. The latter revolve around the central axiom of *psychobiological equivalence* between percepts, experience and thought (Reybrouck, 2013) and address the central question whether there is some lawfulness in the coordinations between sounding stimuli and the responses of music listeners in general. A lot of empirical findings come from studies that reported on *psychophysical dimensions* of music as well as *physiological reactions* that have shown to be their correlates (Peretz, 2001, 2006; Menon and Levitin, 2005; Scherer & Zentner, 2001; van der Zwaag, Westerink & van den Broek, 2011). A distinction should be made, however, between the psychophysics of perception and the psychobiology of the bodily reactions to the sounds. The former suggest a reliable correlation between acoustic signals and their perceptual processing, at least for the levels of sensation and perception; the latter are still subject of ongoing research. Some of them can be subsumed under the sensations of *peak experience*, *flow* and *shivers* or *chills* (Panksepp, 1995) as evidence for particularly strong emotional experiences with music (Gabrielsson & Lindström, 2003). These peak experiences, however, are rather rare and should not be taken as the starting point for a generic comparative perspective on musical emotions. Some broader *vitality effects* or creation of tensions and expectancies may engender some music-specific emotional reactions as well.

Two questions should be addressed further: which *structural features* of the music induce emotions and what are the *underlying mechanisms*? As to the first, dimensions such as musical tempo, timbre, and loudness seem to be important, in addition to increases in perceived complexity, which has been shown to be associated with arousal responses. They seem to be grounded in the dispositional machinery of individual music users and may function as universal cues for the emotional evaluation of auditory stimuli in general. Many questions, however, are still open with regard to their affective and interpretative components, as emotion in music is not a unitary phenomenon. Much more research is still needed in order to trace its underlying mechanisms, though a major contribution to cover a broad range of them has been made already by Juslin & Västfjäll (2008) who present a framework that features six basic mechanisms: brain stem reflexes, evaluative conditioning, emotional contagion, visual imagery, episodic memory, and musical expectancy.

5. EVOLUTIONARY CLAIMS: EMOTIONS AS ADAPTATIONS

Music, as a widespread phenomenon all over the world, can be considered from an *evolutionary* approach, which conceives of it as a universal phenomenon with *adaptive power* (Cross, 2009-2010; Huron, 2003; McDermott & Hauser, 2005; Wallin, Merker & Brown, 2000). As such, it can be approached from different scales of description: the larger evolutionary scale (*phylogeny*) as against the scale of individual human development (*ontogeny*). A lot of evidence comes from developmental (newborns studies and infant-directed speech) (Trehub, 2003) and comparative research (referential emotive vocalizations and separation calls, see above) which have shown that evolution has given emotional sound special time-forms that arise from frequency and amplitude modulation of relatively simple acoustic patterns (Panksepp, 2009-2010).

It has been postulated, further that intact perception and experience of emotion is vital for survival in the social environment. This holds true, especially for the *primary* or *basic emotions* that human beings experience early in life—their listing embraces fear, anger, disgust, surprise, sadness, and happiness—and which reflect the basically innate neural machinery that is required to generate somatic states in response to certain classes of stimuli (Damasio, 1994). As such, they are inherently biased to process signals and to pair them with adaptive somatic responses with the processing proceeding in a preorganised fashion.

The study of the emotional domain is thus related to the evolutionary approach with emotions being considered as *adaptive responses* to be aroused silently in human beings as the product of neural structures that are specialized for their processing. As such, they behave as reflexes in their operation, occurring with rapid onset, through automatic appraisal and with involuntary changes in physiological and behavioral responses (Peretz, 2001). The basic instinct for survival, in fact, is driven by phylogenetically older parts of the nervous system and is embedded in subcortical, instinctive emotional behavior that pre-dates cognitive behavior by hundreds of millions or years (LeDoux, 1998). It involves reactive activity—mostly physiological constants, such as the induction or modification of arousal as well as bodily reactions with a whole bunch of autonomic reactions—, which involves a direct coupling between sensory input and resulting effects. Such “primitive” processing mechanisms, in fact, have considerable adaptive value in providing the organisms with levels of elementary forms of decision making which rely on sets of neural circuits which do the deciding (Damasio, 1994). They can be considered as quasi-universal reactions to sounding music, which are not culture-specific and which may function as universal cues for the emotional evaluation of auditory stimuli in general.

Musically induced emotions, accordingly, can be considered as *reactive behavior* that points into the direction of automatic processing beyond conscious and deliberate control, involving a lot of biological regulation that engages evolutionary older and less developed structures of the brain, and which may have originated as adaptive responses to acoustic input from threatening and nonthreatening sounds (Balkwill & Thompson, 1999).

There are, however, *secondary emotions* as well, which embrace reactions to a broad range of stimuli which are filtered by an interposed voluntary and mindful evaluation and which allow room for variation in the extent and intensity of preset emotional patterns (Damasio, 1994). Human beings, in fact, are wired-in to respond with an emotion to certain features of stimuli in the world, but emotional reactions cannot be exhaustively explained by pre-programmed reaction patterns alone.

6. CONCLUSION AND PERSPECTIVES

Music and emotion studies have recently received some impetus from neurobiological and psychobiological research. The mechanisms behind the actual induction, however, are not yet totally clear. Emotional processing, in fact, holds a hybrid position: it is the place where nature meets nurture with emotive meaning relying both on pre-programmed reactivity that is based on wired-in circuitry for perceptual information pickup (*nature*) and on culturally established mechanisms for information processing and sense-making (*nurture*). It makes sense, therefore, to look for mechanisms behind the inductive power of the music and to relate them with evolutionary claims and the possible adaptive function of music. Especially important here is the distinction between the acoustic and the vehicle mode of listening and the related distinction between the on-line and off-line mode of listening. Much more research, further, is needed in order to investigate the relationship between music-specific or aeshetic emotions and everyday or utilitarian emotions (Scherer & Zentner, 2008). The latter are triggered by the need to adapt to specific situations that are of central signicicance to the individual's interests and well-being. The former are triggered in situations that usually have no obvious material effect on the individual's well-being. Rather than relying on categorical models of emotion by blurring the boundaries between aesthetic and utilitarian emotions we should take care to reflect also the nuanced range of emotive states, that music can induce. As such, there should be a dynamic tension between the "nature" and the "nurture" side of music processing, stressing the role of the musical experience proper.

7. REFERENCES

Balkwill, L.-L. & W.Thompson (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. Music Perception, *17, 43-64.*

Brown, S.; Martinez, M. & Parsons, L. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. Neuroreport, *15(13), 2033-2037.*

Burkhardt, R.W. (2005). Patterns of behavior. *Chicago: University of Chicago Press.*

Clynes, M. (1977). Sentics: The Touch of Emotion. *London: Souvenir Press.*

Clynes, M. & Walker, J. (1986). Music as Time's Measure. Music Perception, *4(1), 85-120.*

Cross, I. (2009-2010). The evolutionary nature of musical meaning. Musicae Scientiae, Special issue, *179-200.*

Damasio, A. (1994). Descartes' Error: Emotion, Reason, and the Human Brain. *New York: HarperCollins.*

Ellis, R. and Thayer, J., (2010). Music and Autonomic Nervous System (Dys)function. Music Perception, *27(4), 317–326.*

Frayer, D.W. & C.Nicolay (2000). Fossil Evidence for the Origin of Speech Sounds. In Wallin, N., B.Merker & S.Brown (Eds.). The Origins of Music *(pp. 271-300). Cambridge, MA - London: The MIT Press.*

Gabrielsson, A., & Lindström Wik, S. (2003). Strong experiences related to music: A descriptive system. Musicae Scientiae, *7, 157-217.*

Godøy, R.I. (1997). Formalization and Epistemology. *Oslo: Scandinavian University Press.*

Hauser, M. & McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. Nature Neuroscience, *6 (7), 663-668.*

Huron, D. (2003). Is Music an Evolutionary Adaptation? In: Peretz, I. & R.Zatorre (Eds.). The Cognitive Neuroscience of Music *(pp. 57-75). Oxford - New York: Oxford University Press.*

Juslin, P.N. & J.Sloboda (Eds.). (2010) Music and Emotion. Theory, Research, Applications. *Oxford: Oxford University Press.*

Juslin, P. & Västfjäll, D. (2008). Emotional responses to music. Behavioral and Brain Sciences, *5, 559-575.*

Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nature Reviews Neuroscience, *15, 170-183.*

Kramer, J. (1988). The Time of Music. New Meanings, New Temporalities, New Listening Strategies. *New York - London: Schirmer Books.*

LeDoux, J.E. (1996). The Emotional Brain: The Mysterious Underpinnings of Emotional Life. *New York: Simon and Schuster.*

McDermott, J. & Hauser, M. (2005). The Origins of Music; Innateness, Uniqueness, and Evolution. Music Perception, 23, 1, 29-59.

Menon, V. & Levitin, D. (2005). The rewards of music listening: Responses and physiological connectivity of the mesolimbic system. Neuroimage, *28(1), 175-184.*

Nagel, F.; Kopiez, R.; Grewe, O. & Altenmüller, E. (2007). EMuJoy: software for continuous measurement of perceived emotions in music. Behavior Research Methods, *39 (2), 283-290.*

Newman, J. D. (2007). Neural circuits underlying crying and cry responding in mammals. Behavioral Brain Research, *182, 155-65.*

Panksepp, J. (1995). The Emotional Sources of "Chills" Induced by Music. Music Perception, *13()2, 171-207.*

Panksepp, J. & Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. Behavioural Processes, *60, 133-155.*

Panksepp, J. (2009-2010). The emotional antecedents to the evolution of music and language. Musicae Scientiae, Special issue 2009-2010, *229-259.*

Peretz, I. (2001). Listen to the brain: a biological perspective on musical emotions. In Juslin, P.N. & J.Sloboda (Eds.). Music and Emotion: Theory and Research *(pp. 105-134). Oxford: Oxford University Press.*

Peretz, I. (2006). The nature of music from a biological perspective. Cognition, *100, 1-32.*

Reybrouck, M. (2005). A Biosemiotic and Ecological Approach to Music Cognition: Event Perception between Auditory Listening and Cognitive Economy. Axiomathes. An International Journal in Ontology and Cognitive Systems, *15(2), 229-266.*

Reybrouck, M. (2013). Musical universals and the axiom of psychobiological equivalence. In J.-L. Leroy (Ed.), Topicality of Musical Universals / Actualité des Universaux musicaux *(pp. 31-44). Paris, France: Editions des Archives Contemporaines.*

Scherer, K.R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, *40, 227–256*

Scherer, K. & Zentner, M. (2008). Music evoked emotions are different – more often aesthetic than utilitarian. Behavioral and Brain Sciences, *5, 595-596.*

Schneck, D. & D.Berger (2010). The Music Effect. Music Physiology and Clinical Applications*. London - Philadelphia: Kingsley Publishers.*

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In Juslin, P. & Sloboda, J. (Eds). Music and emotion: Theory and research *(pp. 393-414). New York, NY, US: Oxford University Press.*

Trainor, L.J. & Schmidt, L.A. (2003). Processing Emotions Induced by Music. In Peretz, I. & R.Zatorre (Eds.). The Cognitive Neuroscience of Music *(pp. 310-324). Oxford - New York: Oxford University Press.*

Trehub, S. (2003). The developmental origins of musicality, Nature neuroscience, *6(7), 669-673.*

van der Zwaag, M; Westerink, J. & van den Broek, E. (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. Musicae Scientiae *15(2) 250–269.*

Wallin, N; Merker, B. & Brown, S. (Eds.) (2000). The Origins of Music. *Cambridge, MA - London: The MIT Press*

USE OF MUSIC FOR EMOTION MODULATION - TESTING THE VALIDITY OF THE IAAM-SCALES WITHIN DIFFERENT GROUPS OF MUSICAL PREFERENCE INDICATES DIFFERENT QUALITATIVE MUSIC PREFERENCE TYPES

Richard Von Georgi[1], Kai Stefan Lothwesen[2]
[1]Institute of Music Science and Music Education - Justus-liebig-university Gießen / International Psychoanalytic University Berlin, Germany, [2]2institute of Music Science and Music Education - University of Bremen, Germany
Time: 09:30

1. Introduction: The use of music in everyday life is a field of growing interest [1-10]. For its measurement the IAAM (Inventory for the measurement of Activation and Arousal modulation by Music) has been constructed and validated since 2002. The IAAM scales are integrated into the neurophysiological model by Gray & McNaughton [11]. The IAAM provides five scales [12]: relaxation, cognitive problem solving, reduction of negative activation, fun stimulation and arousal modulation. IAAM scales are correlated with health [13-15], psychiatric illness [16, 17], music therapy [18], chill-sensations [19], personality, musical preferences [20, 13] and youth problem behavior [21].

2. Aim of the Study: The study examines the validity of the scales by using the MMR (Music in Mood Regulation Scale) by Saarikallio [22]. With respect to Saarikallio's approach specific hypotheses of scale intercorrelation were formulated. Since in previous studies using the IAAM different qualitative types of musical preferences were found [20] these hypotheses were tested within different music preference groups.

3. Method: 150 students (medics) completed the IAAM and the MMR. Musical preference were recorded by using the four preference groups by Rentfrow & Gosling [23]. Simple correlation and cluster analysis were used to examine the covariance structure of the IAAM- and MMR-scales.

4. Results: The reliability of all IAAM-scales lies above 0.80. The cluster analysis supported the formulated hypotheses. With respect to preference groups a clear distinction between the different kinds of negative affect modulation were found solely within the group of intensive & rebellious. On the other hand, persons with a preference for upbeat & conventional music were found to be only differentiating between a general positive and negative emotion modulation by music.

5. Summary: The IAAM-scales are highly reliable. Both the IAAM and the MMR seem to measure comparable constructs. The results confirm the existence of two different qualitative preference types [24]: the integrative-habitual preference type with a sense for applying specific strategies of using music, and the adaptive-situational type which depends more on situational and social factors. All in all, the study generally confirms that the use of music is an learned strategy for emotion modulation.

DIFFERENCE IN AFFECTIVE RESPONSES TO MUSIC BETWEEN CONGENITALLY BLIND AND SIGHTED ADULTS

Hye Young Park[1]
[1]Ewha Womans University, Korea, Republic of
Time: 10:00

The perception and expression of emotions are basic characteristics of music, and listeners and performers communicate their emotions through the music. Most music and emotion studies have focused on the impact of music on the listener; however, there is lack of research regarding emotional responses to music in individuals with congenital blindness. Despite their normal hearing ability, the processing of emotions induced by music may have a different quality in blind adults compared with the sighted. The purpose of this study is to examine the difference of emotional responses to music between congenitally blind and sighted adults. The factors of examination were the congruence of intended emotion of the music; the perceived emotion of the listener and the intensity of such perceived emotions. These values were compared between two groups. A total of 118 participants (47 congenitally blind, 71 sighted adults) listened to twelve 15-second excerpts and reported which emotion they felt among four common emotions, such as happiness, sadness, anger, and fear. They then graded the extent of intensity of the emotion in percentages. Based on previous studies and with reference of reviews of experts, musical stimuli were selected from various films, since film music is designed to induce strong emotional

responses. Results showed that both groups similarly identified the intended emotions of music. However, there was a significant difference in intensity of emotional responses between the groups ($p < .05$) indicating lower mean average in blind group. The intensity of sadness was higher in the congenitally blind than in the sighted while that of happiness and anger was similar for both groups. For fear, the congenitally blind scored a noticeably lower percentage than the sighted. Based on the results of this study, the lack of visual experience among congenitally blind adults may influence the depth and intensity of their emotional responses to music. This suggests that the limited visual input experienced during their developmental period combined with less opportunity to know music as multisensory stimulation are factors to be considered. Future studies should continue to investigate the psychological mechanism of emotional responses to music in individuals with congenital blindness.

Keynote Address 2

SEOUL 11:00- 12:30 Tuesday 5 Aug 2014

LEARNING AND MEMORY PROCESSES IN MUSIC AND DANCE

Catherine J. Stevens
The MARCS Institute and School of Social Sciences & Psychology, University of Western Sydney, Australia

Time: 11:00

Experimental investigations of the development of expectations for music and movement over short and long timeframes are reported. Listeners internalize the regularities of musical environments to which they have been exposed. For example, listeners enculturated to the conventions of Western tonal music have difficulty perceiving and producing asymmetric or uneven meters that occur infrequently in the Western tonal corpus. In a series of experiments, we investigated the hypothesis that adult listeners can learn rhythms implicitly, i.e., without awareness, and can learn culturally unfamiliar rhythms and meters implicitly. Results support the hypotheses although cognitive demand can moderate learning. What about learning movement to music? While there is extensive research on music cognition and growing interest in dance cognition, there are relatively few studies of the learning of dance to music. Sampling participants without formal training in music or dance, we investigated whether the presence of melody, harmony, rhythm, and timbre helps or hinders the learning of a new dance-pop routine. The presence of a rhythm accompaniment during learning provided a significantly greater memory advantage for the recall of the dance-pop steps than full music. The cues to long-term memory (LTM) for dance were scrutinized in another study. The dancers initially recalled the exercises in silence, although each one had been learned to a particular piece of music. Qualitative data indicated that music and dancer movement were important cues to LTM; transitions were sometimes forgotten; images associated with movement recall were kinesthetic, verbal, visual and auditory. Effects of long-term exposure to a particular language background on recognition of musical and spoken items have also been investigated. Accuracy and reaction times of adult participants from tonal (Thai) and non-tonal (Australian English) language backgrounds were recorded as they discriminated Thai and English items that differed in rising/falling contour (speech task), and musical items that differed in rising/falling contour, major/minor interval, and contour plus interval features (music task). The tonal language group was significantly faster and more accurate at discriminating speech items on the basis of pitch contour. The tonal language group was also significantly faster in response to musical contour and intervals, although accuracy was equivalent across language groups. The set of studies reveals that unfamiliar temporal structure can be learned implicitly, temporal structure is sufficient for explicitly learning a dance, and attunement to pitch variation in spoken language can generalize to pitch change in musical materials.

Session4

[4A] Symposium 1

SEOUL 14:00- 16:00 Tuesday 5 Aug 2014

WHAT EMPIRICAL MUSICOLOGY OFFERS EMOTION STUDIES

Discussant: Richard Ashley, Northwestern University, USA

Introduction

The study of music and emotion is an inherently interdisciplinary enterprise. Approaches include controlled experimental work, clinical and case studies, developmental, social psychological, and neurophysiological investigations, among others. The purpose of this symposium is to highlight research on music and emotion that is grounded in empirical musicology, and to discuss a number of recent intersections between the two areas of research. Five papers will investigate the aspects in which emotion studies and musicological questions coincide with one another. Two of the papers feature recent work on newly-curated corpora. The first of these employs a dataset of 19th-century art songs to investigate whether standard prosodic emotional cues common in spoken language are present in the melodies of 19th century German and French art song. The second of these corpus-based studies employs sentiment analysis and probabilistic topic modeling on a dataset of more than 700 opera libretti to investigate the changes in the interaction of character archetypes over time. Another two papers examine the relationship between musical structure and affect. The first of these examines the role of musical interest as an emotional response to music, and discusses how musical structure might influence interest. The second paper is an experimental study that explores the relationship between cues such as modality, timing, and pitch height on the perceived valence and intensity of pieces from Bach's Well-Tempered Clavier. A fifth paper integrates psychological, musicological, and ethological perspectives on music and emotion, and provides musical illustrations of a wide range of plausible affective mechanisms. The

aim of the symposium is to discuss studies in which aspects of empirical musicology and the psychological study of music and emotion overlap, and in which the combined can potentially provide converging evidence with other areas of music psychology.

Interest as an Emotional Response to Music

Elizabeth Hellmuth Margulis
University of Arkansas, USA
Time: 14:00

1. Background: Much research has examined the emotions music can evoke, from sadness to anger to elation. In recent years, Paul Silvia has amassed evidence suggesting that interest qualifies as an emotional response. A casual glance at ordinary listening suggests that something like interest might be an even more common response to music than sadness or elation. Yet since the psychoaesthetics literature of the 1970s, the interest response to music has been little explored.

2. Aims: Adopting the framework for interest as an emotional response provided by Silvia (2006), this paper outlines a way of conceptualizing and testing the notion of interest as an emotional response to music. It reviews recent evidence in the music perception literature for a disassociation between interest and enjoyment. It examines the link between interest and peak experiences of music as chronicled by Gabrielsson (2011). It provides an account of the ways that listener interest can be tracked in a laboratory setting.

3. Main Contribution: By assessing the evidence for the role of interest, this paper opens up new possibilities for conceptualizing and investigating emotional responses to music.

4. Implications: Thinking about interest as an important emotional response to music can illuminate broader questions about aesthetic response, the role of music in everyday life, and the relationship between musical structure and musical affect.

From score to stage: Exploring the relationship between compositional structure and emotional response

Diana Martinez, Lorraine Chuen, Michael Schutz
McMaster Institute for Music and the Mind, Canada
Time: 14:30

1. Background: The communication of emotion in music is a foundational issue in music cognition (Meyer, 1956). Studies involving manipulations of simple melodies as well as novel melodies composed explicitly for research demonstrate that cues such as modality, timing, and pitch height play crucial roles. This project extends that work exploring the interrelationship between these cues in their "natural environment" – unaltered excerpts of J.S. Bach's classic *Well–Tempered Clavier (Book 1)*. This composition serves as an ideal test corpus as it contains 2 pieces in each major and minor key.

2. Aims: We explored the degree to which structural cues predict valance and intensity ratings in each of the *WTC's* 48 pieces. Specifically, we were interested in how pitch height, timing and mode (previously quantified by Poon & Schutz, 2011) predict perceptual responses of emotion to music, as well as the relative strength of each cue.

3. Method: Undergraduate students heard excerpts performed by Friedrich Gulda. They rated valence (1–7) and intensity (1–100) of perceived emotion according to Russell's two–dimensional model of affect (1980). We used a multiple linear regression analysis to develop predictive models of valence and intensity ratings from the structural cues.

4. Results: The three-cue predictor model accounted for 46% ($p<.001$) of the variance in valence ratings and 38% ($p<.001$) of the variance in intensity ratings. of the three cues, timing (i.e. the number of note attacks per-second) was the best predictor of both valence and intensity. Consistent with previous results, modality also played an important role in predicting both dimensional outcomes. Contrary to previous findings in speech and music, pitch height did not correlate significantly with valance in minor key pieces when accounting for timing.

5. Conclusions: Timing best predicted emotion ratings within this corpus, consistent with development research (Dalla Bella et al., 2001), and informing analogous studies exploring parallels in the communication of emotion in speech and music. Most importantly, this rich data set complements previous research using manipulated/novel melodies by exploring the relationships between these cues within the context of literature by one of western music's great composers – J.S. Bach.

Using Sentiment Analysis and Probabilistic Topic Modeling to Examine the Tessitura-Sociability Relationship of Operatic Characters

Daniel Shanahan
University of Virginia, USA
Time: 15:00

1. Background: In a recent study, Shanahan and Huron (in press) examined the effect of tessitura on the perceived sociability, age, and authority in operatic characters. They found that characters with higher tessituras are more likely to be viewed as exhibiting pro-social behavior, whereas characters with lower tessituras are were given significantly lower sociability ratings. This is in accordance with much of the ethological work that suggests pitch height indexes positive and negative social affect (Bolinger, 1975; Morton, 1977;1994). The connection between sociability and tessitura, however, is complicated by the fact that perceived age and authority are also strongly correlated with vocal tessitura in opera. Ideally, a study would be able to provide a more in-depth analysis of the types of interactions

between archetypal operatic characters, which would allow for a more complete investigation into the relationship between pitch height and sociability in the genre.

2. Aims: By conducting a sentiment analysis of a corpus of opera libretti, this study examines the interaction of various character archetypes, and analyzes diachronic changes in the interaction of these archetypes. For example, it examines whether authoritative males exhibit more anti-social behavior toward females who aren't in a position of power, and does this change over time? Following the work conducted by Shanahan and Huron, it's hypothesized that characters with lower-tessituras are significantly will interact more aggressively toward characters with higher tessituras than they would toward characters of similar vocal ranges, ages, or levels of authority. This paper also demonstrates the musicological benefits of both sentiment analysis and topic modeling.

3. Methodology: This paper examines a newly-curated corpus of more than 700 opera libretti, which has encoded each operatic character by perceived age and authority, and employs sentiment analysis derived from many language-specific affective word lists, such as the Berlin Affective Word List (BAWL, which provides affective analysis of German text), AFINN-111 (for English text), FAN (for French text), and the Linguistic Inquiry and Word Count (which allows for affective analysis of Italian Text). It also employs probabilistic topic modeling (specifically, Latent Dirichlet Allocation, or LDA) to better contextualize the interactions between character archetypes.

4. Results: Data is currently being collected.

5. Conclusion: This study demonstrates the benefits of both sentiment analysis and probabilistic topic modeling in musicological research, and further examines the nature of sociability and tessitura in opera. In doing so, it further investigates the role of sound-size symbolism in musical composition.

Integrating Psychological, Musicological and Ethological Perspectives on Music and Emotion

David Huron
Ohio State University, USA
Time: 15:30

ABSTRACT

In attempting to understand how music represents or evokes emotions, different research communities have taken different approaches. A rich psychological literature has arisen, where common topics include affective neuroscience, basic emotions, dimensional approaches, cognitive assessment, and evolutionary-psychological theorizing. Independent of the psychological literature, a musicological tradition exists that has focused on chronicling stylistic conventions intended to convey or represent specific affects in a given cultural context. This approach has documented historically-situated conventions referred to as "topics." Apart from these approaches, ethologists have taken yet another approach to affect. Without access to the subjective experiences of animals, ethologists have instead studied various display behaviors and have developed methods for inferring the motivational states of observers and displayers. It is proposed that each of these three research communities has something to offer in understanding music-related emotion.

1. BACKGROUND

How does music represent or evoke emotions? Attempts to answer this question have appealed to a wide range of mechanisms, from conditioned responses and learned associations, to ethological signals and mirror neuron systems. Apart from the psychological research, issues concerning the representation and evoking of emotion have also been the focus of music scholarship—most notably, musical semiotics and topic theory.

Topic theory has grown over the past three decades beginning with the work of Ratner (1980), Allenbrook (1983), and Agawu (1991). "Topics" are cultural templates that shape particular musical gestures or genres. These include expressions such as the *Hunting Horn* gesture, or the lilting 6/8 rhythm of the *Pastorale*.

In a parallel scholarly track, musical semiotics draws inspiration from the pragmatic philosophy of Charles Sander Peirce (Nattiez, 1987; Lidov, 1999; Tarasti, 2002; Tagg, 2013). Apart from Peirce's influence in the arts, Peirce also inspired central concepts in ethology—the study of animal communication (Tinbergen, 1964; Immelmann & Beer, 1989; Zahavi & Zahavi, 1997; Maynard Smith & Harper, 2003). Both semiotics and ethology make use of central Peircean concepts, such as signal and index. However, there are points of conflict as well as points of convergence between the animal behavior and arts communities.

In this presentation, an integrative framework is proposed and illustrated with musical examples.

2. AIMS

The aim of this presentation is to integrate topic theory, semiotics, and ethological insights, with mainstream psychological views concerning emotion. The purpose is to reconcile some of the points of disagreement, and assemble a check-list of proposed affective and representational mechanisms.

3. MAIN CONTRIBUTION

The main contribution is a more comprehensive inventory of plausible mechanisms for how music represents and evokes affect. The inventory includes mirror systems, signals, indices, cues, learned associations/conditioned responses, symbols, icons, and topics.

4. IMPLICATIONS

The main claims are illustrated using various musical examples.

5. REFERENCES

Agawu, K. (1991). Playing with signs: A semiotic interpretation of classic music. Princeton, NJ: Princeton University Press.

Bradbury, J.W. & Vehrenkamp, S.L. (1998). Principles of animal communication. Sunderland, MA: Sinauer.

Dahl, S., Huron, D., Brod, G., & Altenmüller, E. (2014). Preferred dance tempo: Does sex or body morphology influence how we groove? Journal of New Music Research, DOI: 10.1080/09298215.2014.884144

Ekman, P. (1982). Emotion in the human face. New York: Cambridge University Press.

Frick, R.W. (1985). Communicating emotion: The role of prosodic features. Psychological Bulletin, 97, 412-429.

Hatfield, E., Cacioppo, J.T., & Rapson, R.L. (1994). Emotional Contagion. New York: Cambridge University Press.

Huron D., Dahl, S., & Johnson, R. (2009). Facial expression and vocal pitch height: Evidence of an intermodal association. Empirical Musicology Review, 4, 93-100

Iacoboni, M. (2008). Mirroring people: The new science of how we connect with others. New York: Farrar, Straus and Giroux.

Ilie, G., & Thompson, W. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. Music Perception, 23, 319-329.

Immelmann, K., & Beer, C. (1989). A dictionary of ethology. Harvard University Press.

*Johnstone, R.A. (1997). The evolution of animal signals. In J.R. Krebs & N.B. Davies**Error! Bookmark not defined.** (eds.), Behavioural ecology. (pp. 155-178). Oxford: Oxford University Press.*

Juslin, P.N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin, 129, 770-814.

Katz, J. (1999). How emotions work. Chicago: University of Chicago Press.

Kivy, P. (1990). Music alone: Philosophical reflections on the purely musical experience. Ithaca, NY: Cornell University Press.

Labott, S.M., Martin, R.B., Eason, P.S., & Berkey, E.Y. (1991). Social reactions to the expression of emotion. Cognition and Emotion, 5, 397-419.

Levinson, J. (1997). Music and negative emotion. In J. Robinson (ed.), Music and meaning. (pp. 215-241). Ithaca, NY: Cornell University Press.

Lewis, M. (1995). Self-conscious emotions. American Scientist, 83, 68-78.

Lidov, D. (1999). Elements of semiotics (Semaphores and Signs). London: Palgrave Macmillan.

Marks, I.M. (1987). Fears, phobias, and rituals: Panic, anxiety, and their disorders. Oxford: Oxford University Press

Morton, E.S. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols & J. Ohala (eds.), Sound symbolism. (pp. 348-365). Cambridge: Cambridge University Press.

Nattiez, J.-J. (1987). Music and discourse. Princeton, NJ: Princeton University Press.

Ohala, J. (1980). The acoustic origin of the smile. Journal of the Acoustical Society of America, 68, S33.

Panksepp, J. (1998). Affective neuroscience: The foundations of human and animal emotions. Oxford: Oxford University Press.

Plazak, J. & Huron, D. (2011). The first three seconds: Listener knowledge gained from brief musical excerpts. Musicae Scientiae, 15, 29-44

Ratner, L. (1980). Classical music: Expression, form, and style. New York: Macmillan.

Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. Frontiers in Psychology, 4, 1-33

Schellenberg, G., Corrigall, K., Ladinig, O., & Huron, D. (2012). Changing the tune: listeners like music that expresses a contrasting emotion. Frontiers in Psychology, 3, 1-9. DOI: 10.3389/fpsyg.2012.00574

Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. Psychological Bulletin, 99, 143-165.

Schutz, M., Huron, D., Keeton, K. & Loewer, G. (2008). The happy xylophone: Acoustic affordances restrict an emotional palate. Empirical Musicology Review, 3, 126-135.

Singer, T., Seymour, B., O'Doherty, Stephan, K. E., Kaube, H., Dolan, R. J., et al. (2004). Empathy for pain involves the affective but not sensory components of pain. Science, 303, 1157-1162

Maynard Smith, J.M., & Harper, D. (2003). Animal signals. Oxford: Oxford University Press.

Sobin, C., & Alpert, M. (1999). Emotion in speech: The attributes of fear, anger, sadness, and joy. Journal of Psycholinguistic Research, 28, 347-365.

Sundberg, J. (1987). The science of the singing voice. Dekalb, Illinois: Northern Illinois University Press

Tinbergen. N. (1964). The evolution of signaling devices. In W. Etkin (ed.), Social behavior and organization among vertebrates. (pp.206-230). Chicago: University of Chicago Press.

Taag, P. (2013). Music's meanings. Mass media music Scholar's Press.

Tarasti, E. (2002). Signs of music: A guide to musical semiotics. Mouton de Gryter.

Tiemann, L., & Huron, D. (2011). Beyond happiness and sadness: Affective associations of lyrics with modality and dynamics. Empirical Musicology Review, 6, 147-154.

Tomkins, S.S. (1980). Affect as amplification: some modifications in theory. In: R. Plutchik and H. Kellerman (eds.), Emotion: Theory, research and experience. (pp.141-164). New York: Academic Press.

Voland, E., & Grammer, K. (Eds.) (2003). Evolutionary aesthetics. Heidelberg: Springer Verlag.

Whybrow, P.C. (1997). A mood apart: The thinker's guide to emotion and its disorders. New York: Harper Collins.

De Wied , M., Zillmann, D., & Ordman, V. (1994). The role of empathic distress in the enjoyment of cinematic tragedy. Poetics, 23, 91-106.

Wiley, R.H. (1983). The evolution of communication: information and manipulation. In T.R. Halliday & P.J.B. Slater (eds.), Communication. (pp.82-113). Oxford: Blackwell.

Williams, C.E., & Stevens, K.N. (1972). Emotions and speech: Some acoustic correlates. Journal of the Acoustical Society of America, 52, 1238–1250.

Yano, C.R. (2002). Tears of longing: Nostalgia and the nation in Japanese popular song. Cambridge, MA: Harvard University Press

Zahavi, A., & Zahavi, A. (1997). The handicap principle: A Missing Piece of Darwin's Puzzle. Oxford: Oxford University Press.

Zentner, M., Grandjean, D., & Scherer, K.R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. Emotion, 8, 494-521

[4B] Music and Movement 3

KYOTO 14:00- 16:00 Tuesday 5 Aug 2014

THE WALKING WITH MUSIC IN AN URBAN ENVIRONMENT

Marek Franek[1], Leon Van Noorden[2], Lukas Rezny[1]
[1]University of Hradec Králové, Faculty of Informatics and Management, Czech Republic, [2]University of Gent, Institute for Psychoacoustics and Electronic Music, Belgium

Time: 14:00

The aim of the study was to investigate effects of music listening while walking on walking speed and emotional changes. Selection of music was based on the concept of motivational music characterized by fast tempo and strong rhythm, which stimulates fast and intensive bodily movements (Karageorghis, Terry, and Lane, 1999). 120 undergraduate students (60 females) took part in the experiment. Participants were walking on a 2 km long circuit through various environments (park, quiet streets, or streets with a heavy traffic) in the town of Hradec Králové in the Czech Republic. They listened to either motivational or non-motivational music (nice music with no strong urge to move) during they walk. The walking speed was measured in 16 sections with the length about 50 m. The participants were asked to complete the Positive and Negative Affect Schedule (PANAS) before and after the walk. Moreover, they completed a personality questionnaire based on the Big-Five personality model. Results showed that motivational music made the participants to walk faster than the participants listened to the non-motivational music (1.7 m/s vs. 1.4 m/s). Further, it was found that the tendency to synchronize with the music was low. In motivational music condition, positive emotions significantly increased in females after the walk, whilst the level of negative emotions significantly decreased in both genders after the walk. In non-motivational music condition, only negative emotions significantly decreased after the walk in females. Further, in motivational music condition it was found negative correlation between walking speed and neuroticism, while in non-motivational music condition it was found negative correlation between walking speed and extraversion indicating that personality features effect reaction on music listening while walk. The results demonstrated effects of different types of music on behavioral and emotional reactions during the walk, as well as the effect of personality.

KINAESTHETIC DIMENSIONS IN THE AESTHETIC PERCEPTION AND EXPERIENCE OF MUSIC

Jin Hyun Kim[1]
[1]Humboldt University of Berlin, Germany

Time: 14:30

1. Background: In recent music research focusing on embodied perception and cognition, music has been conceived of as multimodal. Accordingly, the number of empirical studies directing focus toward embodied, gestural features of musical performance and bodily reactions to music has increased. However, the role of kinaesthesia in music perception, especially in the aesthetic perception and experience of music, has received little scholarly attention.

2. Aims: This paper discusses theoretical considerations concerning the role of kinaesthesia in the aesthetic perception and experience of music.

3. Main contribution: Psychological and philosophical aesthetic theorists such as Theodor Lipps, John Dewey and Gregory Currie suggest that the coupling of action and perception underlies the aesthetic experience. In particular, kinaesthetic simulation of motor action is characterized as a basic mechanism of aesthetic experience. Likewise, in recent neurocognitive research, such motor simulation is increasingly assumed to guide the observation of action executed by others and characterized as a significant (neural) process pointing to embodied cognition. In musicology and music pedagogy of the fin de siècle, expressive forms of music were regarded as coupled with the experience of music. The concept of expressive forms of music as dynamically (co)shaped finds empirical evidence in recent studies in developmental psychology on infants' multimodal behavior, which Daniel N. Stern refers to as 'forms of vitality.' These forms of vitality as manifested in music prove to be experiential features rather than ostensibly objective musical features. Hence it can be assumed that measured physical features can be mutually related to reported subjective qualities of the aesthetic experience of music. The present author's case studies on musicians' aesthetic experiences during performance show that verbal expressions used in performers' reports are largely related to kinaesthetic dimensions of experience—kinaesthetic images in cases of musical flow, and kinaesthetic sensations in cases where that flow is disturbed.

4. Implications: Kinaesthetic images related to the processes of motor simulation underlying the aesthetic perception and experience of music can hardly be inferred from either behavioral studies or measured physiological reactions; therefore, the kinaesthetic images guiding musicians in their aesthetic experience deserve careful empirical investigation, followed by discussion on methodological considerations.

TEMPORAL STRUCTURES EMBODIED THROUGH DANCE—INTERACTIONS OF PERCEPTION, COGNITION, AND ATTEMPTED ESCAPE FROM THE PERIODIC TEMPORAL PERCEPTION WINDOW

David Mosher[1]
[1]*University of Massachusetts, Amherst, USA*
Time: 15:00

The periodic temporal perception window is a frame of time in which we as humans have been shown to perceive temporal regularities. This window of time ranges from between one hundred milliseconds to about six seconds. We may outwardly demonstrate this periodic attending by means of entrainment. This process (also called synchronization or attunement), refers to the fact that the human response to periodic input is to become phase-locked to the external stimuli. Through this process, part of our biological activity synchronizes with periodic environmental occurrences. The fact that we are synchronizing biological activity, often in hierarchic manner, is indicative of our initial perception of levels of periodicity inherent in metric frameworks, as well as our ability to continue to predict future periodic onsets within the recurring pattern. Entrainment can involve not only our perception and attention, but also encourage bodily motion. This includes all kinds of motor activity correlated with musical motion including clapping our hands, tapping our toes, or even dancing. In this manner, perception, cognition, and motor activity are linked in our directed attention to musical meter. This paper further examines the relationships between the perception window and our outwardly visible entrainment through dance. A dance step performance analysis of the pavane is presented with the purpose of examining the relationships of its execution to the temporal span of the perception window. Efforts to embody escape from the window are considered in representations of 'the fleeting moment' as well as a butoh training manual in the art of slow movement. Through these examples the following related questions are explored: What do our own movements show us as outward evidence of our inward cognitive and perceptual metric perceptions? How and in what ways is our bodily motion in time shaped and grounded by our periodic temporal perception window (PTPW)? Is it possible to embody an escape from the window? What could various dances reveal through their embodied relationships to the window? What could these embodied relationships to the window show us about possible ideological frameworks of congruent music and dance structures?

"IN SEARCH OF LOST TIME": CONNECTIONS BETWEEN THEORIES OF HEARING AND PERFORMANCE DISCOURSE

Youn Kim[1]
[1]*The University of Hong Kong, Hong Kong, China*
Time: 15:30

ABSTRACT

The ideas of shaping music and its embodied nature, in conjunction with the attention given to the trajectory of musical motion, are being briskly advanced both in music theory and psychology. This paper examines how the precedent notion of music as process and endeavors to trace it emerged in the late 19th and early 20th centuries in various attempts to find the "lost time" in physiology, aesthetics, music theory, and performance discourse. As a case study in the history of music psychology, it aims to establish the connections between various domains of perception, movement, and music. The phrase "lost time" (*temps perdu*) is mostly known as the title of the 1913 masterpiece novel of Marcel Proust. Its earlier origin is, however, found in scientific literature. About 60 years earlier, one of the greatest natural scientists of the 19th century, Hermann von Helmholtz (1851), used precisely the same expression in his letter to the French *Académie des Science*. Describing and securing the priority of his nerve-muscle experiments on frogs, he used the phrase to describe the time lag between stimulus and the nervous response. Contrary to the previous presumption of the instantaneous transmission of sensory information, Helmholtz demonstrated a finite and measurable nerve conduction time, and thus made a milestone in neurophysiology. This finding also carries a critical significance for the history of psychology. As Boring (1957) writes, it dramatically advertises that "mind is not ineffable but a proper subject for experimental control and observation."

Less widely acknowledged is the conceptual significance of lost time for contemporaneous music-psychological, and by extension, music-theoretical discourse. The phenomenon illustrates a new kind of psychological time in which the lines between past and present are blurred. Hence, it conforms to the continuity of perceptual experience, which is also illustrated by the then widely discussed physiological notion of after-sensations. Referring the sonic impression lingering even after the stopping of the actual sound, "after-sounding" (*Nachklingen*), as afterimage in vision, led to the shift away from referentiality to the external world and toward the subjectivization of hearing (Crary 1990).

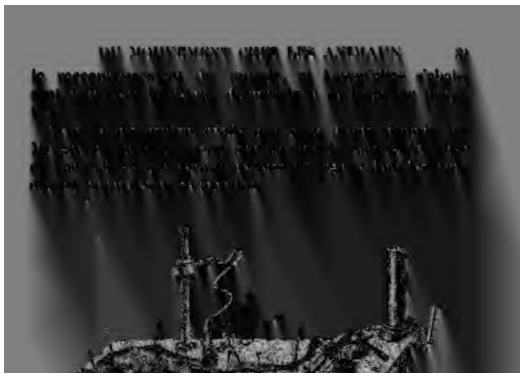


Figure 1: "Marey's Myograph." Source: Marey, É.-J. (1873). La Machine animale: Locomotion terrestre et aesrienne. Paris: Masson.

These discussions of temporality of hearing are consolidated with the issue of motion. An interesting nexus is found in the works of Étienne-Jules Marey, the physiologist and photographer who propagated Helmholz's ideas to the French scientific community. Possibly, the expression "lost time" might have reached Proust through this route. Marey was particularly concerned with temporality of the moving body (1868, 1873, 1894). Not only did he discuss the issues of the pianist's gesture and musical notation, his ideas also, directly and indirectly, influenced early-20th-century writings on music performance. In his path-breaking objective and scientific study on piano technique, Otto Ortmann (1929) adopted Marey's photographic method to trace the hand and arm movements of the pianist.

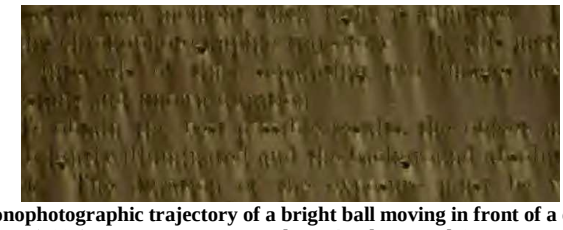


Figure 2: "Simple Trajectory and chronophotographic trajectory of a bright ball moving in front of a dark background." Source: Marey, É.-J. (1895). *Movement*. New York: D. Appleton and Company.

Our musical perception is shaped by our awareness of the action producing the sonic events. Given such close connections between bodily gestures and music as uninterrupted experience, the early-20th-century "energeticist" music theory that stresses the dynamic quality of music is not only speculative and abstract but also entangled with performativity. Ernst Kurth's "melody as motion," Heinrich Schenker's "piano singing" and, in particular, their discussions of portamento at the piano bespeak what the pianist Theodor Leschetizky described once: "There is more rhythm between the notes than in the notes themselves."

These discussions of the time lag in perception and hearing and musical gaps gave rise to the question of the agency of dynamic shaping. Musical scores were conceptualized as "scripts" in both editions and music analysis. Simultaneously, the multilayered nature of the musical mind was revealed to include what may be referred to as "auditory," or more precisely, the "musical unconscious," which is comparable to Walter Benjamin's "optical unconscious."

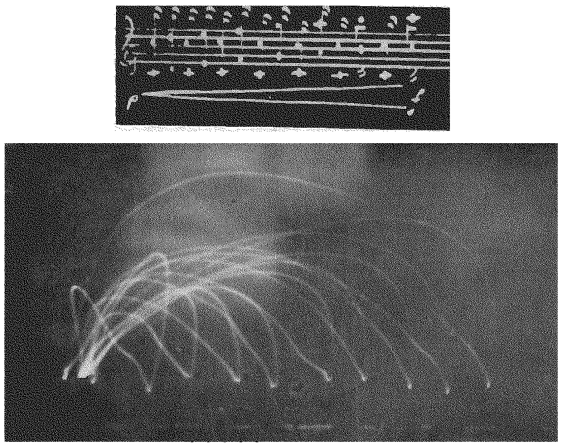


Figure 3: Photograph of hand movement in the octave figure with a crescendo. Source: Ortmann, O. (1929). *The Physiological Mechanics of Piano Technique*. New York: E. P. Dutton.

REFERENCES

Baily, J. (1985). Music structure and human movement. In P. Howell, I. Cross & R.West (Eds.), Musical structure and cognition *(237-258). London: Academic Press.*

Benjamin, W. (1999). Little history of photography. In M. W. Jennings (Ed.), Walter Benjamin: selected writings, volume 2 *(507-530). Cambridge: The Belknap Press of Harvard UP.*

Boring, E. G. (1957). A history of experimental psychology. Englewood Cliffs, NJ: Prentice-Hall.

Crary, J. (1990). Techniques of the observer: on vision and modernity in the nineteenth century. *Cambridge, MA: MIT Press*

Helmholtz H. (1851). "Deuxième Note sur la vitesse de propagation de l'agent nerveux," Comptes rendus hebdomadaires des séances de l'Académie des Sciences, 33, *262-65.*

Marey, É.-J. (1868). Du mouvement das les fonctions de la vie. *Paris: Baillière.*

Marey, É.-J. (1873). La Machine animale: Locomotion terrestre et aesrienne. *Paris: Masson.*

Marey, É.-J. (1895). Movement. *New York: D. Appleton and Company.*

Newcomb, E. (1967). Leschetizky as I knew him. *New York: Da Capo Press.*

Ortmann, O. (1929). The Physiological mechanics of piano technique*:* an experimental study of the nature of muscular action as used in piano playing, and of the effects thereof upon the piano key and the piano note. *New York: E. P. Dutton.*

Kurth, E. (1931). Musikpsychologie. *Berlin: Max Hesses.*

Schenker, H., & Esser, H. (2000). The art of performance. *New York: Oxford University Press.*

Schenker, H., & Rothgeb, J. (1987). Counterpoint: A translation of Kontrapunkt. *New York: Schirmer Books.*

Stern, D. N. (2010). Forms of vitality. *New York: Oxford University Press.*

[4C] Music and Health

MONTREAL 14:00- 16:00 Tuesday 5 Aug 2014

ACTIVE AGEING: THE SYMBIOSIS OF MUSIC AND HEALTH IN THIRD AGE RELATIONSHIPS

Jill Morgan[1], Raymond Macdonald[1], Paul Flowers[2], Dorothy Miell[1]
[1]The University of Edinburgh, United Kingdom, [2]Glasgow Caledonain University, United Kingdom

Time: 14:00

Varied academic accounts exist of the emotional responses experienced through engagement with music. MacDonald et al (2002) explore how music 'as a channel of communication' provides a 'means by which people can share emotions, intentions and meanings'. However there has been less focus on the older members of society and in a study by Pickles (2003) into music and the third age a plea is made 'for better understanding of the musical needs and opportunities for this age group'. With the population increase in those aged over sixty and contemporary studies indicating a positive correlation between long term partnerships and lifespan longevity (Jaffe et al 2011), there is a need for research into the value of music and its role within couple relationships in retirement. The aim of this research is to explore the role of musical experiences in the wellbeing of marriages/partnerships in a group of five co-habiting couples who are retired and in good health, seeking to offer new insights into the importance of music in a key area of health, identity and social relationships. The work aims to access participants' in depth accounts within the following domains: personal emotional responses to music, music through the lifespan, perceived uses of music within a relationship, the role of music in the health of couple relationships, the nature of music as a communication tool and the formation of a joint identity. An idiographic methodology was adopted using Interpretative Phenomenological Analysis. In-depth semi-structured interviews were conducted individually, with each participant offering meaningful perspectives on the phenomenon being investigated; data was recorded and analysed. Preliminary findings suggest the utilisation of music as an instrument for joint identity and as a future leisure tool when health and mobility issues become problematic. Conclusions will be drawn from the full results and discussed within the wider context of music and health research.

DOES SINGING FACILITATE SOCIAL BONDING?

Gunter Kreutz[1]
[1]Carl Von Ossietzky University Oldenburg, Germany

Time: 14:30

Psychobiological effects of amateur choral singing were studied in a naturalistic controlled within-subjects trial. A mixed group of novice and experienced singers (N = 21) filled out brief questionnaires of psychological wellbeing and gave samples of saliva at the beginning of two rehearsal sessions and thirty minutes later. The singing condition included warm-up vocal exercises and repertoire pieces (e.g., "California Dreaming"). In the chatting condition, dyads of participants talked to each other about recent positive life experiences. Within-subjects, repeated measures analysis of variance (ANOVA) on self-reported and physiological measures (salivary oxytocin, cortisol, and dehydroepiandrosteron (DHEA)) revealed significant Time X Condition interactions for psychological wellbeing and oxytocin. Comparisons of mean scores showed patterns of changes favouring singing over chatting. There were no significant interactions for cortisol, DHEA as well as for the cortisol-DHEA-ratio. These results suggest that singing enhances individual psychological wellbeing as well as induces a socio-biological bonding response.

A QUALITATIVE INVESTIGATION OF DEVELOPMENTAL DIFFERENCES IN THE FUNCTIONS OF MUSIC LISTENING AND ITS ROLE IN EVERYDAY WELL-BEING.

Jenny Groarke[1], Michael Hogan[1]
[1]School of Psychology, National University of Ireland, Galway, Ireland
Time: 15:00

1. Background: A review of the literature reveals a wide range of music listening functions, with listening to music for mood regulation being reported as the most important function. Current models of well-being highlight not only positive emotions, but also meaning and engagement as orientations to happiness.

2. Aims: The current study sought to use collective intelligence methods to address how music listening may facilitate subjective and eudaimonic well-being in younger and older adults. The study involved a novel application of the collective intelligence methodology, Interactive Management (IM). IM is useful for generating a priori models grounded in participant's consensus-based logic.

3. Method: Two IM sessions were conducted, one with younger adults aged 22-29 years (N = 11: 6 male, 5 female) and one with older adults aged 60-72 (N = 9: 5 male, 4 female). Participants generated ideas in response to the stimulus question 'why do you listen to music?'. Each participant then voted for five ideas, which they believed, were most significant for increasing well-being. The eight highest ranked functions of music listening were entered into Interpretive Structural Modelling (ISM) software, and relations between pairs of ideas were discussed, 'Does function A significantly enhance function B?'. The group discussed each relational question, and an enhancement structure was generated.

4. Results: Overall, there was a large degree of convergence in music listening behaviours across both groups. Analysis of structural models revealed age group differences in the highest-ranking functions of music listening. The older adults in this study emphasised more eudaimonic functions of music listening (e.g. Transcendence, social connection), whereas the younger adults valued music's role in affect regulation. Differences were also observed in the structuring of interdependencies between functions.

5. Conclusions: Our findings support previous studies demonstrating that music listening, especially for affect regulation, appears to be more common in adolescence. This may reflect their greater experience of negative affect relative to older adults. These findings suggest that cohort differences in music listening may be due to the contrasting demands of everyday life for different age groups, and that both younger and older adults are using music adaptively to meet these demands.

THE ROLE OF MUSIC IN THE EXPERIENCE OF FLOW IN BALLET DANCERS

Clorinda Panebianco-Warrens[1]
[1]University of Pretoria, South Africa
Time: 15:30

This interpretative phenomenological study explores professional ballet dancers' subjective experiences of flow in a performance context. It further aims to gain a deeper understanding of the role of music in flow. It was assumed that ballet dancers' experience of flow would be similar to flow experiences by performers in other domains, such as sport, but that music would play a more important role in flow than has been reported to date. Seventeen ballet dancers shared their experiences in separate semi-structured interview sessions. The interviews took place during the production of Don Quixote and Camille respectively, so that the dancers could draw on recent performance experience in their recollection of flow. The interviews were transcribed and analysed both deductively and inductively. The results showed that ballet dancers share similarities in flow with flow in other domains according to Csikszentmihalyi's dimensions of flow. The three dimensions that stood out as most predictive of flow in this study were the merging of action and awareness, autotelic experience and loss of self-consciousness. It was found that experience, music and the level of musical skills also seemed to play a prominent role in flow experienced by ballet dancers. The study argues that processing music places an extra demand on dancers' concentration and considers possible modes of processing information.

[4D] Music Performance, Expression and Objective Analysis

LIÈGE 14:00- 16:00 Tuesday 5 Aug 2014

EVALUATION OF A BOTTOM-UP MODEL OF MUSICAL SIGHT-SINGING: THE CASE OF MODUS NOVUS

Ricardo Goldemberg[1], Philip Fine[1]
[1]University of Buckingham, United Kingdom
Time: 14:00

ABSTRACT

Believing that the abundance of knowledge in regards to language can provide valuable insights into the mechanisms that underline the reading of music, this paper adopts a comparative approach that corroborates the proposition of a bottom-up approach to the acquisition of sight-singing, especially in the beginning stages of learning. The logic behind *Modus Novus*, a workbook created by Edlund Lars for dealing with the difficulties of reading 20th century music, is assessed and provides the underlying principles of the learning of sight-singing based on a model which emphasizes the systematic study of musical intervals as the smallest unit of musical code in both tonal and non-tonal contexts.

1. BACKGROUND

The ability to sight-read music through vocal intonation has been highly esteemed in the musical world for centuries. Prominent musicians consider it as an indication of intelligent musical thought since it is only when musicians become able to read music actively and independently that it becomes possible to exploit music in a unique way, constituted by a world of music literacy.

Sight-singing skills, when fully developed, are strongly associated with higher-order cognitive processes such as aural imagery, i.e. the ability to imagine sounds in silence. Benward (1980) says that musicians often describe this ability as a definition-defying sense of musical awareness, that is commonly characterized by the "hearing eye [and] seeing ear" (p. vii).

Sight-singing, beyond its intrinsic value as a dynamic and interactive language of communication that allows users to share specific content, is a process of immense educational potential in developing people´s musical understanding and music literacy.

2. THE CORRELATION BETWEEN MUSIC AND LANGUAGE

Only with the advent of cognitive musicology – at the end of the 20th century – has research in the area of music and language become more consistent. The approach based on new paradigms accompanied by innovative methods and concepts has led to a deeper understanding of aspects that were previously subject only to speculative analyses.

The hypothesis of structural sharing was suggested in the seminal works of Patel (1998 onwards), where contradictory observations derived from neuropsychology and neuroimaging are reconciled by suggesting that "linguistic and musical syntactic processing engage different cognitive operations, but rely on a common set of neural resources for processes of structural integration in working memory" (Patel, 1998, p. 27).

According to Patel, the "shared syntactic integration resources hypothesis" or SSIRH, distinguishes between two types of cognitive networks in language and music cognition – representational networks and resource networks. As explained by Fiveash and Pammer (2012), "representational networks can be viewed as domain-specific knowledge areas, where information about music and language are held in long-term memory … Resource networks, on the other hand, are the domain general processes by which this task-specific information is accessed and used in working memory" (p. 2). Patel´s hypothesis suggests that the overlap in music and language occurs at the level of resource networks.

In support of the SSIRH there is already a growing body of evidence, most of which is derived from experiments where syntactic manipulations of a linguistic and musical nature are crossed in search of interactive effects that may be indicators of shared processing (Slevc et al, 2012, p. 3414).

Besides the possibility of structural links, verbal and musical reading share procedural mechanisms that are not exclusive to both domains. Jackendoff (2009) verifies that specific cognitive capacities like "large-scale memory, combinatoriality, expectation, imitation, innovation, and joint action" (p. 197) are not unique to language and music, but are general capabilities that also extend to other domains.

Thus, music and language share nonexclusive cognitive processes in which "the use of discrete building blocks and the generation of higher order structures through combinatorial rules is a major point of similarity between music and language" (Brown, 2000, p. 273), leading to the assertion that "all formal differences between language and music are a consequence of differences in their fundamental building blocks (arbitrary pairings of sound and meaning in the case of language; pitch-classes and pitch-class combinations in the case of music). In all other respects, language and music are identical" (Pesetsky and Katz, 2009, p. 3).

3. STRATEGIES OF VERBAL READING

Today, reading is understood as a complex skill involving the integration of distinct and complementary components that range from discriminating symbols to comprehending text. These two poles delineate the reference axes of all theoretical reading models. Based on these differences, teachers have adopted their own educational approaches providing frequent polarization between bottom-up or top-down strategies.

The bottom-up strategy comes from a mechanistic view of language where it is understood that the process of reading begins with the printed text, the identification of letters, syllables and words, gradually progressing to comprehending the written text. In this case, the first task of the reader is to decode the written symbols via the phonological route. From this conversion the information is gradually refined, and through lexical and syntactic clues available in the text, meaning is built.

In contrast to the bottom-up model, the approach to reading from a top-down stand point is based on the principle that reading is a process that begins in the mind of the reader who, from a few clues, speculates and constructs meaning in a continuous procedure of confirmation and reassessment of hypotheses. This is a "psycholinguistic guessing game" (Goodman, 1967) in that the reader is no longer completely dependent on the written text and, according to Barnett (1989), "uses general knowledge of the world or of particular text components to make intelligent guesses about what might come next in the text [and] samples only enough of the text to confirm or reject these guesses" (as cited in Lally, 1998, p. 268).

However, trying to explain the reading process as purely bottom-up or top-down is limited and touches on a number of conceptual difficulties, giving rise to a more balanced and integrative perspective in which distinct processes act together. In this case the reader, while processing information by identifying and decoding words, reconstructs the meaning of the text based on their knowledge of the world. This is a context in which bottom-up and top-down strategies operate in a way that mutually facilitate each other and "the linguistic knowledge coming from various sources (spelling, syntax and semantics) interact in the reading process" (Zimmer, 2006, p. 51).

4. L2 READING ACQUISITION

While striving for integrated strategies is a frequent goal in the teaching of reading, almost all researchers agree that developing vocabulary is critical in understanding and interpreting texts. Accordingly, the L1 (first language) reader brings with him or her the baggage of proficient knowledge of language which immediately allows the emphasis of using top-down strategies based on inferential processes, while L2 (second language) readers do not have similar resources leading to the need to "hold in the bottom"(Eskey, 1988, pp. 96-97).

In fact, the existence of linguistic barriers in learning L2 reading goes beyond the evident limitations of vocabulary. Grabe (1991) explains that “even if a reader has good command of L2 syntax, it is unlikely he/she will be familiar with the practical knowledge, culturally established understanding through social interaction among speakers of a foreign language” (as cited in Zimmer, 2006, p. 52).

Due to these limitations, Eskey (1988) stresses his concern that promoting high-level strategies can be excessive. He believes that hierarchically lower skills such as decoding and rapid and accurate identification of grammatical patterns are not mere obstacles to be overcome in the "guessing game", as suggested by supporters of primarily top-down models, but skills that must be effectively consolidated to avoid excessive inferential demands, decreasing the guess work in comprehending text.

In this way, "teachers must not lose sight of the fact that language is a problem in reading a second language, and that even the influence of plausible guesses are no substitute for accurate decodification" (Eskey, 1988, p. 97).

5. MUSIC AS L2 – BOTTOM-UP PROCESSES IN EARLY READING

Access to pre-established schemas in the case of music, representing *a priori* knowledge, may be possible due to acculturation processes to which we are all subject. Virtually everyone acquires a basic repertoire of melodies and a strong predisposition to tonality early in life, constituting a preliminary reference in the formal learning of music. This process of "tonal acculturation manifests itself in the form of schemas or patterns of perceptual activities from which the individual learns the sound form he is subject to. Notes, melodies, chords are used within precise contexts, taking advantage of the tonal schemas encoded by use and execution of music of the Western world for more than 300 years"(Grossi, 1994, notes).

The emphasis on tonality is usual in teaching sight-singing, regardless of the approach adopted. Movable systems explore the matter more explicitly, although the materials used are practically the same, by the usual means of collections of melodies and musical passages arranged in order of difficulty. Still, it should be noted that although the inferential character of tonality facilitates the process, there are huge individual differences.

The fact is that in learning to sight-sing, it is frequently difficult to access previously developed structural patterns much like in learning L2 reading, where the existence of inherent language barriers limits the inferential process of recognition. While the use of top-down reading strategies is not completely excluded in light of our acculturation in a tonal environment, their support bases are considerably tenuous in early learning. Frequently, hearing schemas are only partially acquired before formal learning takes place, even leading to situations where learning to sight-sing must begin practically from a clean slate.

Therefore, even in music, it is also recommended to emphasize bottom-up processes in learning to read. It is only in so far as a repertoire of recurrent musical-sound ideas are developed, that the reading process takes on an interactive nature, with the increasing use of complementary top-down strategies.

In these circumstances, there is clear merit for teaching initiatives that focus on systematically acquiring the skill to articulate minimum units of code, whether linguistic or musical. From a learning point of view, it is a process based on the construction of consistent associations between fundamental elements of sound and symbols, which is identified with the teaching of phonics in verbal reading.

The goal of phonics teaching is to enable beginner readers to handle, manipulate and create through understanding of the relationships between the basic units of written and spoken language. As a method, it proposes planned, systematic work with the phonemic-graphemic relationships, as well as the underlying cognitive skills for the establishment of these relationships, that is, phonological awareness (Souza and Rodrigues, 2009, p. 20).

Similarly, “musical phonics" is a plan of work with basic units of music where the interval of sung or written music corresponds by association to the spoken phoneme or written grapheme of language respectively. This analogy is supported by Fletcher (1957) when stating that both spoken language and sung music consist of a linear series of sounds produced vocally, where the difference only lies in the quality of the spoken sound, or in the pitch and rhythm of the music.

Although the idea of a one-one correspondence between the structural elements of the two domains is only partial and certainly limited, there is clearly a similarity in terms of processing since "in order to use the code at all, the would-be 'encoder' or 'decoder' must develop skill in a mode of hearing such that the continuous sound-pattern is apprehended as a series of units corresponding to units available in the code" (Fletcher, 1957, p. 77).

6. THE CASE OF MODUS NOVUS

The dissolution of tonality in the 20th century brought new challenges. Consequently, traditional sight-singing methods, used effectively up until then, had difficulties in dealing with the complexities of atonal music. The focus on tonality, the basis of solid teaching, simply ceased to be relevant. If before, using inferential resources in sight-singing was already relatively limited, especially during the beginning stages of learning, it had now become completely unrealistic, even compromising the skills of proficient readers.

Therefore, the bottom-up approach to sight-singing became practically the only strategy available with virtually no top-down support. This new method took shape with a proposition developed by Lars Edlund, a Swedish professor of the Royal Academy of Music in Stockholm, in his textbook entitled *Modus Novus: Studies in Reading Atonal Melodies* published in 1963.

Edlund says (referring to 20th century music), that considering the lack of principles of logical structure that could serve as a method of aural training, the musical interval can be seen as an "atonal" figure capable of generating music discourses whose meaning is far beyond the simple ratio between two notes. Although the technique proposed is reductionist in nature, as it deals with a limited set of intervals each time, the author clearly states that "the student's command (visual and aural) of the theory of intervals in the absolute sense of the word … is here merely a *pre-requisite* for the further study of what I would like to call the 'aural study of the musical patterns' " (1963, pp. 13-14).

From a technical standpoint, Edlund takes particular care to stress that the domain of individual intervals alone is no guarantee of accuracy in reading atonal melodies, and that "the most important thing now is to practice *combinations of intervals* that will break the bonds of the major/minor interpretation of each individual interval" (1963, p. 13). Therefore, sight-singing training is always conducted through the use of a series of musical intervals, avoiding tonal contextualization. Each chapter of *Modus Novus* presents a new interval, always in the company of other intervals previously introduced in varying combinations and in progressive order of difficulty, consisting of extensive practice of atonal melodies composed by known intervals, adding the new interval which is the focus of the chapter.

7. INTERVALLIC SIGHT-SINGING

In the 50 years since the publication of *Modus Novus*, a few additional works were published which increased the scope of its application to include less sophisticated musical contexts. The new proposals basically maintained the same rationale of gradual study of musical intervals presented in series and atonal contexts, but simplified it by adding additional steps when introducing musical intervals, elaborating preparatory exercises, and eliminating difficulties of rhythm.

What is particularly relevant is that the benefits of these exercises are not limited to 20th century atonal music, but have clear applications for the tonal music that precedes it. The understanding by certain authors is that training with intervals "is particularly helpful in dealing with music that is extremely chromatic, tonally ambiguous or rapidly modulating" (Prosser 2010, cover), while providing a broad scope when one "must learn and memorize the sound of each interval as *it is of itself*, usable in any musical context" (Prosser, 2010, p. 6).

Also, Adler (1997), in a textbook that clearly presents the principles of tonality, states that the "ability to sing all intervals within any musical context, tonal or non-tonal, is the goal of this text" (p. xi). In this case, the author organizes a pedagogical collection of original and composed melodies, many of which are highly modulating and chromatic, explaining that "they should be practiced carefully and sung at first purely by interval; after several repetitions, when each piece is integrated into the student's musical psyche and ear, the tonal scheme will be apparent" (p xi.).

In particular, this last statement is consistent with the notion that reading melodies with low inferential content, as well as sight-reading in its early stages, requires the use of a predominantly bottom-up approach, and it is only when one gradually forms a reference repertoire of musical patterns, tonal or non-tonal, that a top-down approach is used with more frequency.

Even in learning to read in the mother tongue, with fewer inherent limitations, the generative nature of the bottom-up process leads to progressive "lexicalization" where different processing strategies gradually become complementary. In these circumstances, "each successful decoding encounter with an unfamiliar word provides an opportunity to acquire the word-specific orthographic information that is the foundation of skilled word recognition" (Share, 1995, p. 155).

8. FINAL COMMENTS

From a psychological point of view, the practice of sight-singing is a complex, high-level process that can be considered an act of rebuilding from visual-perceptual stimuli, involving significant interaction between conceptual knowledge and the expectations of the reader. Although clearly susceptible to training, individual differences are extensive, and the strategies used by fluent readers vary considerably depending on factors such as the type of music, familiarity with the genre, self-confidence, knowledge of music theory and melodic awareness.

In this context, "musical phonics" should not be understood as one isolated or specific method of teaching in lieu of others; rather, it is a poorly disseminated structuralist approach with immense teaching potential in that it has a heavy emphasis on consolidating the foundations. The understanding is that learning to read comes gradually in a process that originates with decodification followed by comprehension. It assumes that the reader well trained in basic skills will have cognitive resources freed, being able to fully decipher the notation and comprehend it better.

Hopefully, pedagogical efforts founded prominently on basic skills will help "the development of sight-singing skills (that) may bridge the purely external rendition of the music via the hand [of the performer] to the developing representational power of the mind behind the musical voice" (Scripp, 1995, p. 26), leading ultimately to what Dunnett (1994) calls "an intuitive aural grasp of pitch, interval, rhythm, harmony and musical structure - the 'natural' musicality" (as cited in McNeil, 2000, p.73). This, in itself, would represent an enormous achievement in the world of music.

9. REFERENCES

Adler, S. (1997). Sight-singing: Pitch, interval, rhythm. *New York, NY: W.W. Norton.*
Benward, B. (1980). Sightsinging complete. *Dubuque, IA: Wm. C. Brown.*

Brown, S. (2000). The "musilanguage" model of music evolution. In N.L. Wallin (Ed.), The origins of music *(pp. 3-24). Cambridge, MA: MIT Press.*

Edlund, L. (1963). Modus novus: Studies in reading atonal melodies. *Stockholm: Edition Wilhelm Hansen.*

Eskey, D. E. (1988). Holding in the bottom: An interactive approach to the language problems of second language readers. In P.L Carrell, J. Devine & D.E. Eskey (Eds.), Interactive approaches to second language reading *(pp. 93-100). Cambridge: Cambridge University Press.*

Fiveash, A., & Pammer, K. (2012). Music and language: Do they draw on similar syntactic working memory resources? Psychology of Music, 42*(2), 1-20.*

Fletcher, S. (1957). Music reading as a code-learning problem. Journal of Music Theory, 1*(1), 76-96.*

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. Journal of the Reading Specialist, 6, *126-135.*

Grossi, C. (1994). A abordagem não-tonal na aprendizagem musical: Um relato de pesquisa. Cadernos de Estudo: Educação Musical, 4*(5), 52-59.*

Jackendoff, R. (2009). Parallels and non-parallels between language and music. Music Perception, 26*(3), 195-204.*

Lally, C. (1998). The application of first language reading models to second language study: A recent historical perspective. Reading Horizons, 38*(4), 67-77.*

McNeil, A. F. (2000). Aural skills and the performing musician: Function, training and assessment. *(Doctoral dissertation, University of Huddersfield). Retrieved from http://eprints.hud.ac.uk/4749/1/270448.pdf*

Patel, A. D. (1998). Syntactic processing in language and music: Different cognitive operations, similar neural resources. Music Perception, 16*(1), 27-42.*

Pesetsky, D., & Katz, J. (2009). The identity theses for language and music. *Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, MA. Retrieved from http://ling.auf.net/lingbuzz/000959/v1.pdf*

Prosser, S. (2010). Intervallic ear training for musicians. *Rottenburg: Advance Music.*

Scripp, L. R. (1995). The development of skill in reading music. *(Doctoral dissertation, Harvard University). Available from ProQuest Dissertations and Theses database. (UMI No. 9534633)*

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non *of reading acquisition.* Cognition, 55*(2), 151-218.*

Slevc, L. R., Reitman, J. G., & and Okada, B. M (2013). Syntax in music and language: The role of cognitive control. In Cooperative minds: Social interaction and group dynamics. *Paper presented at the Proceedings of the 35th Annual Meeting of the Cognitive Science Society, Berlin, Germany (pp. 3414-3419). Austin, TX: Cognitive Science Society.*

Souza, A. C., & Rodrigues, C. (2009). Alfabetizar e letrar: Reflexões com base no método fônico. *Retrieved from http://alfabetizacaoeletramneto.blogspot.co.uk/2009/12/alfabetizar-e-letrar-reflexoes-com-base.html*

Zimmer, M. C. (2006). O processamento da leitura em língua materna e em língua estrangeira: Uma abordagem conexionista. Revista Signo, *3, 49-64.*

EIGENPERFORMANCE: METHOD TO DESCRIBE THE PROFICIENCY OF PIANO PERFORMANCE

Masanobu Miura[1]
[1]*Faculty of Science and Technology, Ryukoku University, Japan*
Time: 14:30

ABSTRACT

Analysis of musical performances in terms of proficiency or skill has been widely conducted but no salient method to describe the proficiency of piano have not yet proposed. Aim of present study it to develop a method to describe the proficiency of piano performance, by introducing a principle component analysis (PCA), named "Eigenperformance". The proposed Eigenperformance is a method to take a look what occurs on a mass of recorded data and is realized by obtaining the principle components (PCs) for obtained musical data. Here, MIDI recordings are analyzed in order to obtain the PCs for onset, velocity, and duration of MIDI. Then the eigenvectors of the samples are independently obtained for onset, velocity and duration, and each sample is described as a vector with members of PCs. On the other hand the proficiency of the sample was rated by experts of piano, and the average of the rated scores is used as the subjective scores of the samples. The samples are classified into several classes by the *k*-means algorithm by observing the value of members on the vector, and then the proficiency scores are compared among clusters. The Eigenperformance of piano performance of a simple one-octave and Czerny 40's etude are obtained, which are obtained for all samples and for limited number of performances for each class. It clarifies the tendency of each class, describing the common feature of each class on performance. Concretely, those with high proficiency show stable and gradual changes in performance whereas those for poor proficiency show unstable and sudden changes, remarkably for onset and MIDI-velocity. The duration does not have so much change, since it doesn't provide remarkable perception in case of piano performance (attenuating tones in particular). The proposed Eigenperformance is introduced as a method to describe the tendency of piano performance, and moreover it is hopefully employed in analysis of instrumental performances.

1. BACKGROUND

Performance evaluation for the proficiency of piano performance has been intensively studied, where the piano performance is modeled as a combination of tendency curve and deviation from it (Morita et al., 2010; Nonogaki et al., 2012). The method draws a global tendency of the performance, by using a spline curve, so that the parameters are possible to extract. The method is by and large meaningful since it could achieve to realize an automatic evaluation of the performance. However, the criteria on the proficiency have been still unclear, although the clarification of the criteria on the performance proficiency will be contribute to practical situation of piano

performances. In order to clarify the criteria of the performance proficiency, several studies have been conducted (Shaffer et al, 1987; Sundberg, 1980; Repp, 1992), most of them are for the clarification of skilled player's performances, not for the comparison between skilled and unskilled. Thus, reports for the evaluation criteria by seeing the difference between skilled and unskilled are not familiar in this field.

Therefore, this study draws the global tendency of current performance as previous study did and clarifies the criteria of proficiency evaluation by evaluating the parameters derived from the global curve. Target performance of this study is a scale performance with ascending and descending within one octave comprised by 15 notes. Figure 1 shows an example of the task. The samples of the performances are used from previously used in our study (Miura et al., 2010) that contains 336 samples for the 12 keys with 14 players by right or left hand. Ten of the players are university's students and four are experts of piano. The subjective scores for the samples are rated from 1 to ten (ten is the best) and the averages among the four experts are dealt with.

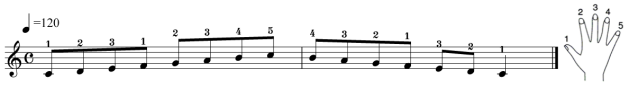


Figure 1: Example of sheet music and fingering number

2. METHOD

2.1. Calculation of spline curve

The global tendency of performances is drawn by a spline curve. Unlike other curves such as n-dimensional polynomial curve or moving averages, the spline curve needs several points to be passed and by allocating the points to the centers of the note clusters categorized by the crossings or turnings on fingerings, the spline curve can deal with the fingerings on the curve, where the inadequate deviations of the notes based on the crossings or turnings are less evaluated compared to the center of clustered notes. Figure 2 shows the outlines of calculation of spline curve.

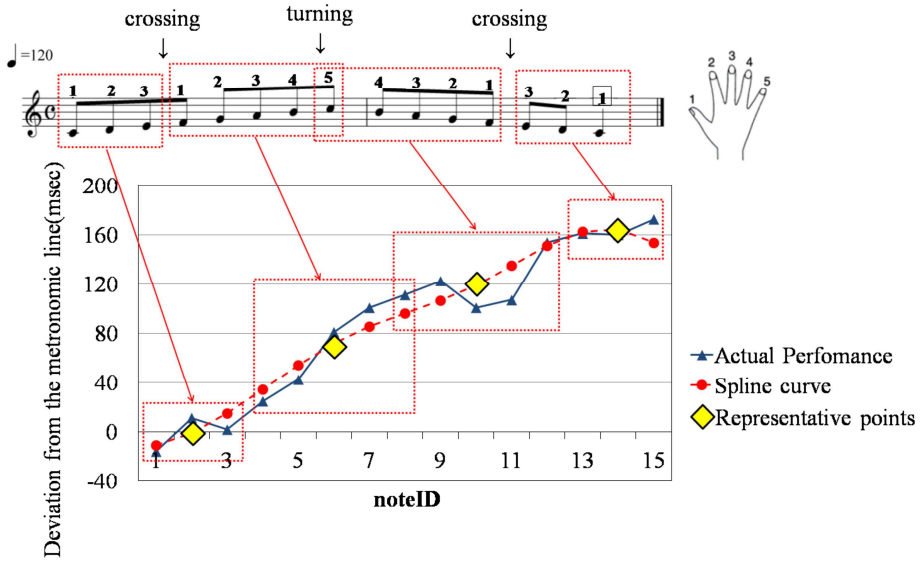


Figure 2: Method to calculate spline curve.

2.2 INVESTIGATION OF EVALUATION CRITERIA

2.2.1 Feature parameters

Table 1 shows the list of feature parameters used in this study, where *i* represents each of onset timing (*i*=t), MIDI-velocity (*i*=v) or duration (*i*=d). Each of timing, MIDI-velocity and durations represents the difference from standards for each (250 msec for timing and durations, and the mean MIDI-velocity for the mean of current velocity). The P_0-P_4 are used in previous study (Miura et al., 2012) and P_5 and P_6 are newly added. P_5 represents the mean score of subjective evaluation score for the nearest seven samples by observing the shape of spline curves. Here why the P5 is employed is to grasp evaluation for the spline curve, but it is not directly obtained from experts so instead of this we employ the mean of the nearest seven samples' scores. P_6 represents the deviation of the difference between recorded and spline curve.

2.2.2*k*-means clustering

To confirm the possibility to represent performance proficiency by the parameters listed in table 1, unsupervised clustering was carried out where each sample is represented by a set of 21 parameters, for seven parameters * three features(onset, velocity and duration). The number of sampled to be classified is 336, and the number of the clusters are four.

2.2.3 Eigenperformances

The Eigenperformances is eigenvectors for plenty of performance patterns. Previously the Eigenperformance is employed on an automatic composition of base part in popular music (Abe et al., 2012), which clarifies the effectiveness of the Eigenperformance to compute the naturalness of the generated patterns. In contrast to the study, the curves categorized to each of the four clusters are independently calculated to obtain Eigenperformance for the spline curves in the clusters. Thus the analysis of the shape of curves on each cluster is conducted. The 90% cumulative contribution ratio of the results of PCA is used on the present analysis.

Table 1: Feature parameters used in this study.

*Pi*0	standard deviation of subtracting standard from recording
*Pi*1	rms deviation of subtracting standard from recorded from spline curve
*Pi*2	range of spline curve
*Pi*3	difference in spline curve between adjacent notes
*Pi*4	sum of spline curve from standard
*Pi*5	Mean of evaluation scores for the nearest seven samples in terms of the shape of spline curve
*Pi*6	Standard deviation of differences between recorded and spline curve

3. RESULT

3.1 Validity of feature parameters

3.1.1 By proficiency estimation

So as to confirm the validity of the feature parameters on the possibility to describe, we conduct proficiency estimation, likewise the previous study did (Miura et al., 2012), by using the 21 parameters under the leave-one-out method. The flow of the proficiency estimation is shown in Figure 3. The result of the proficiency estimation is that R^2 shows .50 and .45 (n=336) when using 21 parameters and 15 parameters, respectively. Moreover, the 18 parameters including the 15 previous parameters and P_5 shows .47and that with the 15 and P_6 shows .47. The result when using only six parameters of P_5 and P_6 shows .35. Therefore, the effectiveness of the newly added parameters was confirmed.

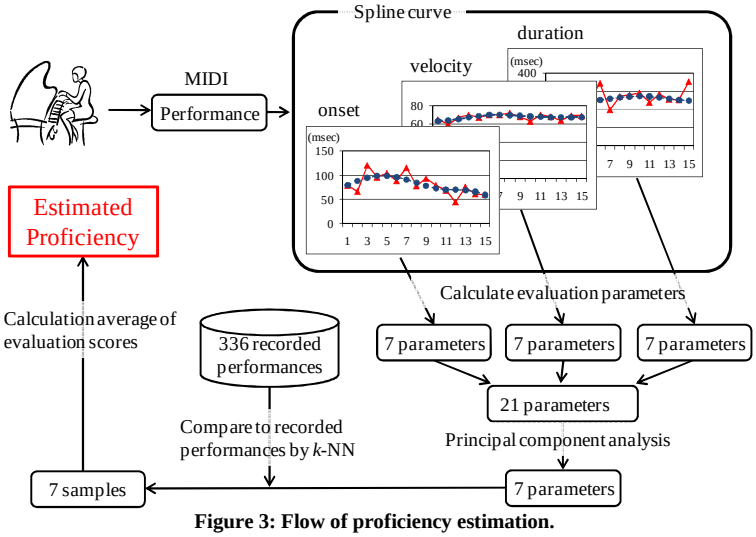


Figure 3: Flow of proficiency estimation.

3.1.2 By unsupervised clustering

The result of unsupervised clustering gives four clusters. The 366 samples are categorized into the four clusters and then labeled in the order of average mean of the subjective scores. Then the four clusters are compared in terms of the subjective scores and then confirmed that all the clusters have significant differences (p<.05, Holm) among them. Therefore, I labeled the clusters into "Excellent(n=158)", "Good(n=121)", "Poor(n=37)" and "Worst(n=20)". The clusters are shown in figure 4, where abscissa represents clusters and ordinate represents subjective scores included in each cluster. Tick represents 95% CI. Then each parameters for each cluster is shown in figure 5, where the value of each parameter is normalized to avg.=0 and SD=1. Tick represents 95% CI.

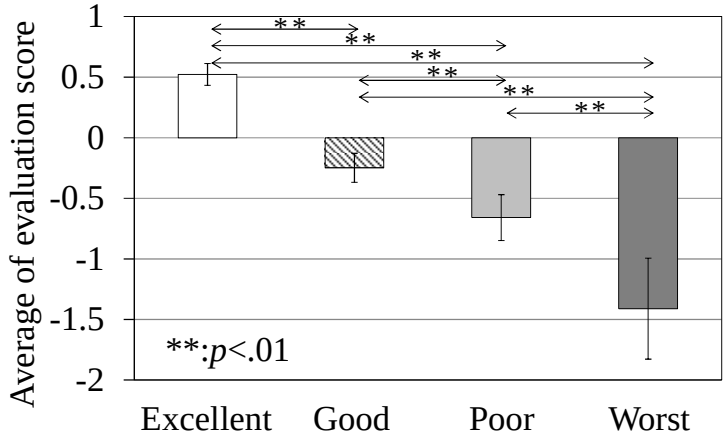


Figure 4: Result of unsupervised clustering and the subjective scores with 95% CI.

3.1.3 DISCUSSION

Figure 5 shows some aspects of the proficient performances. Certain Parameters shows no overlapping to other clusters, for example Pv_0, being thought as one of the specific aspects of the proficient performances. By observing such parameters, we can roughly draw what is the proficient performance from the parameters.

For onset timing, the proficient performances is stable (rarely deviate suddenly) and have less deviations from the spline curve. One of the possible criteria for onset timing is that "Gradual change or those without sudden changes may establish a proficient performance."

For MIDI-velocity, the proficient performances shows stable (, or without any sudden changes) and deviation from spline curve is not so much. Thus, one of the possible criteria for MIDI-velocity is "Playing without any salient change of velocity may be evaluated as proficient".

For duration, the proficient performances show stable changes but not salient change in sum. Apparently possible criteria for the proficient performance is "Playing with smoothed change of duration without any salient change is likely to be evaluated as proficient". Apparently the highest evaluated performance has almost consistent durations, or almost the durations of the note value in the sheet music.

In short, common features on the proficient performances are stableness and smoothed changes. Therefore, aspects of proficient performances are clarified. Moreover, the 95% CI for P_{t3} , P_{t5} , P_{v3} , P_{v5} and P_{d3} of the Excellent cluster show little variability, showing the common feature of the performances of the Excellent cluster.

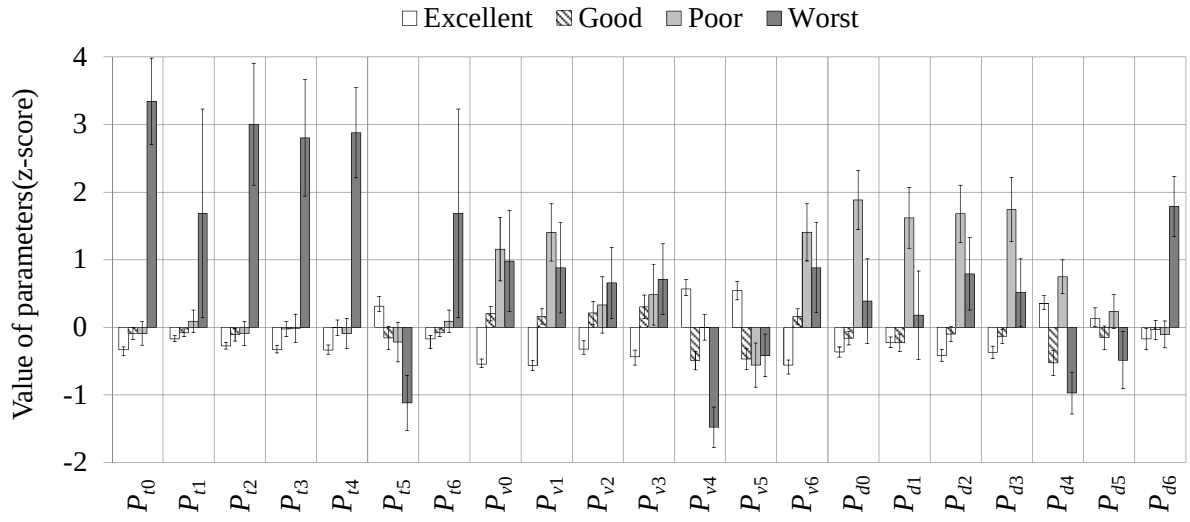


Figure 5: Feature parameters for each cluster with their 95% CI.

3.2 Validity of feature parameters

3.2.1 Calculation of Eigenperformances

Figures 6, 7 and 8 shows the Eigenperformances for the spline curves of the piano scale performances, where the cumulative contribution ratio of PCA is 90%. Therefore, 1st and 2nd principle components (PCs) are for onset timing, only 1st PC are for velocity, and 1st are for durations.

For onset timing, figure 6 shows that the proficient performances (Excellent, in this case) show linear changes. For example, the 1st and 2nd PCs show almost linear change, thus the proficient performances show smoothed changes. Those in Good and Poor show a bit changing patterns. On Worst, the changes are quite large compared to others, thus the badly evaluated performances show several changes on the onset timing.

For the MIDI-velocity, figure 7 shows a gradual change on velocity. For example, PC on Excellent has no salient changes and linear changes as well when comparing with Good and Poor. On the other hand, the PC shows large change on Worst.

For durations, figure 8 shows that the 1st PC shows no salient change on Excellent and Good, implying the proficient performances show no salient changes on duration but being likely to have a constant length. On the other hand the Poor and Worst show large changes and unstable changes.

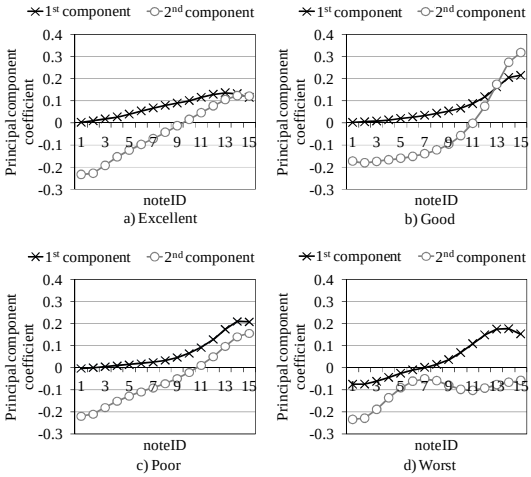


Figure 6: Eigenperformances for onset timing.

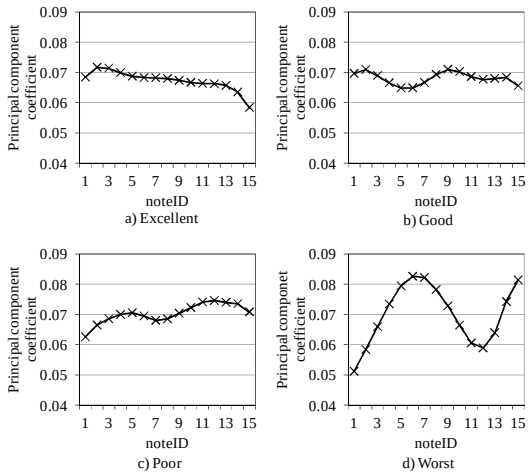


Figure 7: Eigenperformances for MIDI-velocity.

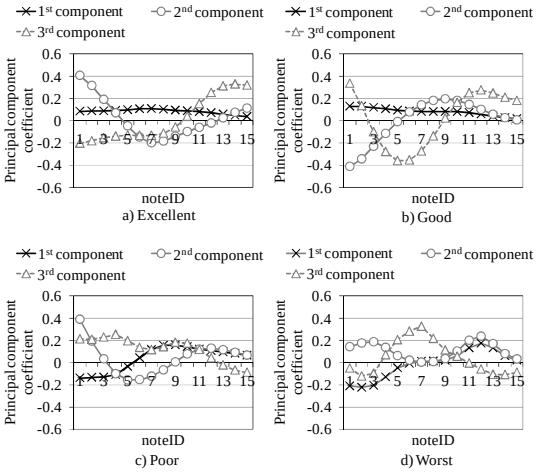


Figure 8: Eigenperformances for duration.

3.2.2 Discussion

Certain aspects of proficient performances clarified here are clarified. For onset timing, gradual change and fewer differences from spline curve is likely to be evaluated as proficient. For MIDI-velocity, gradual change (without any sudden changes) is likely to be evaluated as proficient. For duration, the sloop is almost flat and to be in line to abscissa and has almost a constant length is likely to be evaluated as proficient.

So as to grasp the actual deviations of recorded from spline curve, figure 9 illustrates the amount of differences of recorded performance from spline, where the ordinate represents the normalized average of the differences. As can be seen from figure 9, the proficient performances show little differences from spline curve.

Finally, the Eigenperformance show that the performances closer to the spline curve have a tendency to be evaluated as proficient. In other words, the proficient performances do not include any sudden changes and have a gradual change likewise spline curve has. Contrarily, the less proficient the performances become, the more sudden changes and salient differences the performances have on their global tendency. Therefore, the proficient performances have a spline-curve-like shape on onset timing, MIDI-velocity and durations at all.

Figures 10, 11 and 12 illustrates the example of recorded performances and their spline curves. As can be seen from the figures, the findings described above are intuitively confirmed as true. Therefore, the present study found out the possibility to describe the nature of proficiency by using the Eigenperformance.

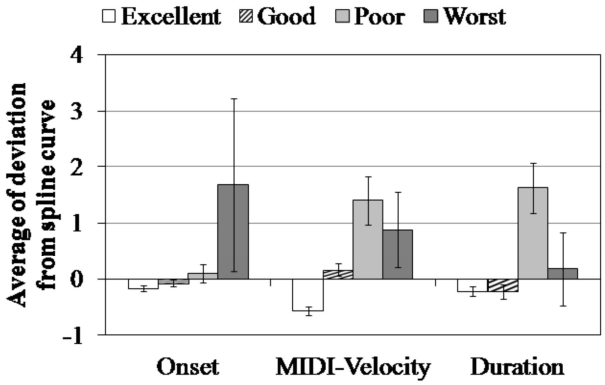


Figure 9: Normalized deviations from spline curves with 95% CI.

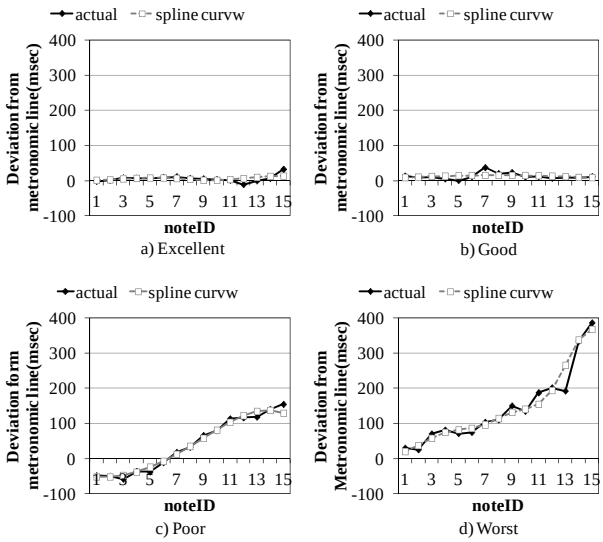


Figure 10: Example of onset timing.

4. CONCLUSION

The present study conducts an unsupervised clustering for the piano scale performances using the *k*-means algorithm to four clusters. The obtained clusters show no overlaps among clusters in terms of the evaluation scores, therefore the clusters are confirmed to be associated with the proficiency and the feature parameters used on the clustering are confirmed as effective for representing the proficiency on the performances. By comparing the value of feature parameters, the tendency of the proficient performances is obtained. Eigenperformance is calculated, to grasp the tendency of the performance . The Eigenperformances for the classified clusters unveil the tendency of the spline curves of performances. The result of *k*-means and Eigenperformance has a consistent result as follows;

Proficient performances have 1) gradual or smoothed changes, 2) no sudden changes, and 3) spline-curve-like changes.

Future works are to conduct this analysis to other performances such as Czerny piano pieces, and to discuss it to the more aesthetic performances like Chopin Etude or so.

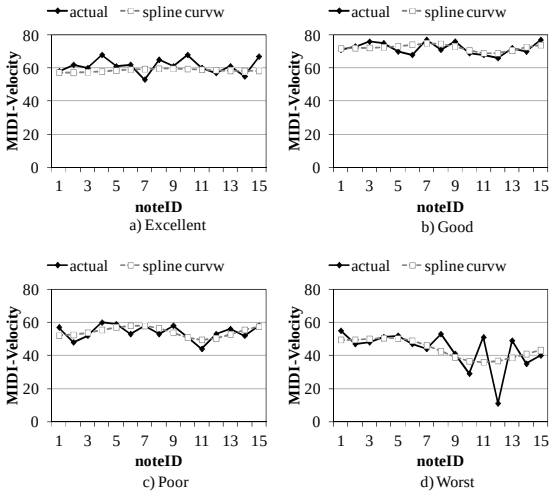


Figure 11: Example of MIDI-velocity.

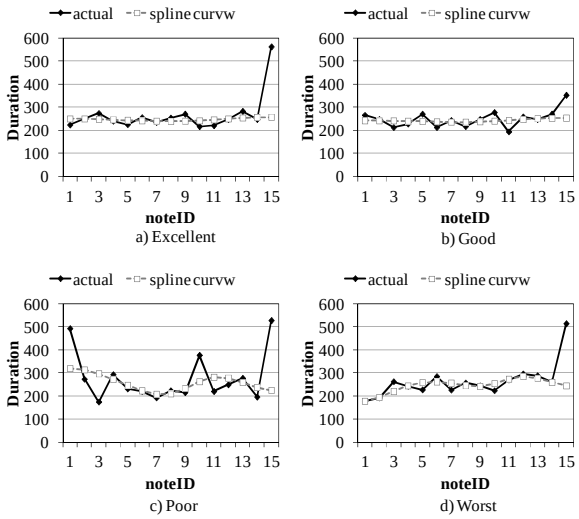


Figure 12: Example of duration.

5. REFERENCES

Abe, Y., et al., *(2012) " Automatic arrangement for the bass guitar in popular music using principle component analysis," Acoustical Science and Technology, 33, pp.229-238.*

Nonogaki, A., et al., *(2012) "Evaluation parameters for proficiency estimation of piano based on tendency of moderate performance," Proc. of 12th International Conference on Music Perception and Cognition, pp.728-737, Thessaloniki.*

Miura, M., et al., (2012) "Automatic evaluation of performance proficiency of one-octave major scales using the piano," J. Acoust. Soc. Jpn, 66, 5, pp.203-212 (in Japanese).

Morita, S., et al., *(2010) "A feature exaggeration system for basic piano performance with extracted performance parameters," Acoustical Science & Technology, 31, pp.296-299.*

Repp, B. H., (1992) "Diversity commonality on music performance: An analysis of timing microstructure in Shumann's Träumerei", Journal of the Acoustical Society of America, 92, pp.2546-2568.

Shaffer, L.H., et al., *(1987) "The interpretive component in musical performance, " In Action and Perception in Rhythm and Music (Gabrielsson, A., ed.), Royal Swedish Academy of Music , pp. 139–152.*

Sundberg, J., et al., *(1980) "On the anatomy of retard: A study of timing in music," Journal of the Acoustical Society of America, 68, pp.772-779.*

THE IMMANENT EMOTION OF A MUSICAL SCORE: AN EXPLORATORY STUDY

Erica Bisesi[1], Richard Parncutt[1], Sylvain Caron[2], Caroline Traube[2]
[1]University of Graz, Austria, [2]University of Montreal, Canada

Time: 15:00

ABSTRACT

By focusing on notated music and separating immanent emotion (latent in the score) from performed emotion (contributed by the individual performer), we are attempting to model immanent emotion as a linear combination of simple predictors for valence and arousal. We calculated a series of relevant measures - such as strength of tonal center, dissonance, and average pitch variation - from a set of musical scores partly selected from the piano repertoire and partly rearranged by one of the author, in order to reproduce different musical structures by varying each aspect one at a time. Performances corresponding to 24 different structures (recorded on a MIDI grand piano Disklavier in 3 different expressive realizations - deadpan, moderately expressive and emotional, for a total of 72 performances) are randomly presented to musicians and non-musicians, who are asked to rate them according to both dimensional and categorical models of emotions. Preliminary results are consistent with hypotheses that positive immanent valence is associated with major or clear tonality, and immanent arousal with tempo and melodic pitch range. A preliminary investigation indicated also that, within an intermediate range of tempi, stimuli corresponding to unclear tonality and atonality, as well as to a high degree of dissonance, are generally associated with feelings like uncertainty, anxiety, fear, and associations like dark, gloomy, insecurity, sense of death.

1. BACKGROUND

How does musical emotion depend on musical structure? Clynes (1977) related the emotional quality of a music event to the shapes of loudness and tempo contours. Sloboda (1991) related structural elements as sequences and new and unexpected harmonies to emotional responses such as shivers and tears. Juslin (2001) related basic emotions (happiness, anger) to the musical surface (tempo, dynamics, articulation); Juslin et al. (2002) presented a computational model of expression in music performance, where generative structure rules, emotional expression and movement principles are integrated in a common picture (GERM model). Zbikowski (2010) explored the correlation between remarkable musical passages (as dissonances and modulations) and emotion changes.

On the other side, research in experimental aesthetics has been greatly aided by the advent of the Hevner adjective list (Hevner, 1935; 1936). After its formulation, the Hevner checklist - consisting of clusters of words grouped by similarity of meaning, arranged in a circle so that adjacent clusters are more similar in meaning than more distant clusters - has been progressively revised by improving its consistency across and within each cluster (Farnsworth, 1954; 1969), and by updating the checklist with a wider spread of emotional ranges (Schubert, 2003).

In a separate study, we are extending existing categorical models for emotions, by exploring qualitative relationships and quantitative correlations between categories and subcategories of music structure, emotion and free associations (Bodinger, Bisesi and Parncutt, 2014).

2. AIMS

This study attempts to clarify the question of the relationship between structure and emotion by first separating emotion in notated music into immanent emotion (emotion that is somehow latent in the score, e.g. the heroic character of Beethoven 3rd symphony or the gloomy character of Liszt Lugubre Gondola) and performed emotion (emotion that is contributed by the individual performer). We explore the relationship between musical structure and immanent emotion by proposing a model of immanent emotion as a linear combination of simple predictors for valence and arousal, and by relating basic musical structures to specific categories for emotion.

3. METHODS

3.1. Participants

18 participants (9 musicians and 9 non-musicians) recruited by email.

3.2. Materials

We have calculated a series of relevant measures from a set of musical scores, such as strength of tonal center (based on frequency of occurrence of scale steps), dissonance (frequency of occurrence of pitch class sets and root ambiguity), average pitch variation, and overall tempo. On this basis, we selected 72 stimuli according to the following prescriptions. First, we selected 4 pieces corresponding to 4 different tonal strengths: 1 piece in a clear major tonality (Field Nocturne No.1 in E flat major), 1 piece in a clear minor tonality (Saint-Saëns Pavane in D minor from the opera Étienne Marcel), 1 piece in an uncertain tonality (Widor, Après la fête Op.71 No.5), and 1 atonal piece (Ligeti Étude No.5, Arc-en-ciel).

Second, these 4 pieces have been modified by one of the authors (who has expertise in music theory and composition), in order to reproduce the different structures listed above. Specifically, for each of the 4 pieces we included 2 different levels of harmonic dissonance (by modifying some of the chords), 2 different melodic ranges (by shifting some pitches without changing the harmonic function of corresponding chords), 2 different performing tempi (fast and slow), and 3 different performances for each of the previous stimuli (1 deadpan, 1 moderately expressive performance focusing on local variations of tempo and dynamics - as suggested by the analysis of structural aspects such as phrase segmentation and accents, and 1 “emotional” performance). These three specific expressive strategies are motivated by previous studies on music performance and emotion (van Zijl et al., 2012). Due to the low level of familiarity of the pieces, we expect that all the modifications in the structure and in the performance (with respect to the original composers’ intentions) will not negatively affect participants’ ratings.

All the performances have been recorded by one of the author (who is a professional pianist and has experience as researcher in music performance and expression) on a MIDI grand piano Disklavier in both MIDI and audio formats.

3.3. Procedure

9 musicians and 9 non-musicians listen to the series of recorded performances in random order, and perform two tasks: (i) they rate the valence and arousal of each piece, and (ii) they select the three words from a modified version Hevner Adjective Clock (cfr. Bodinger, Bisesi and Parncutt, 2014), which they think best describe the piece, providing also three additional words of their choice. Then they associate to each word a rating from 1 to 3 (following its subjective importance).

To run the test (in the participants’ mother language), we developed special computer interfaces (based on Psychopy and Java).

The duration of each stimulus is approximately 30 sec (depending on the performing tempo), and the whole experiment lasts about 90 min.

This study aims to focus on immanent emotion, ignoring for the moment the effect of performed emotion. This is the reason because we are considering three different renditions of each piece: by thrating all the performance parameters (such as variations in tempo, dynamics and articulation) as covariant of the other variables, and given the possibility to extract such parameters from MIDI data, their effect on the results will be estimated and discussed.

All the freely provided words that are not included in the current version of the Hevner Adjective Clock will be used to extend the model.

In the end, we discuss relationships and correlations between measures for musical structure (as listed above), emotions and (if any) free associations.

4. PRELIMINARY RESULTS AND DISCUSSION

Preliminary results are consistent with hypotheses that positive immanent valence is associated with major or clear tonality, and immanent arousal with tempo and melodic pitch range.

To identify uncertain or ambiguous tonality and/or atonality, as well as to describe unfamiliar contexts for consonance/dissonance, categorical models of emotions look more appropriate than dimensional models.

A preliminary investigation indicated that, within an intermediate range of tempi, stimuli corresponding to unclear tonality and atonality, and/or higher degrees of dissonance are generally associated with feelings like uncertainty, anxiety, fear, and associations like dark, gloomy, insecurity, sense of death.

Further details are in preparation and will be presented at the conference.

5. ACKNOWLEDGEMENTS

This research is supported by the FWF Stand-Alone Project P 24336-G21 “Expression, emotion and imagery in music performance”. We kindly thank Dieter Kleinrath (KUG, Graz) for his support in the implementation of the Java computer interface used in this study.

6. REFERENCES

Bisesi, E., Bodinger, M., & Parncutt, R. (2013). Listeners’ informal vocabulary for emotions and free associations in piano music. Abstract accepted for spoken presentation at ICME 2013, Jyväskylä, Finland, 2013.

Bodinger, M., Bisesi, E., & Parncutt, R. (2014). Listeners’ informal vocabulary for structure, expression, emotions and associations in piano music. Abstract accepted for spoken presentation at ICMPC-APSCOM 2014, Seoul, Korea, 2014.

Clynes, M. (1977). Sentics: The touch of emotions. New York: Doubleday Anchor.

Farnsworth, P. R. (1954). A study of the Hevner adjective list. Journal of Aesthetics and Art Criticism, 13, *97-103.*

Farnsworth, P.R. (1969). The social psychology of music (2nd Ed.). Ames, I.A.: State University Press.

Hevner, K. (1935). The affective character of the major and minor modes in music. American Journal of Psychology, 47, *103-118.*

Hevner, K. (1936). Experimental studies of the elements of expression in music. American Journal of Psychology, 48, *246-268.*

Juslin, P. N. (2001). Communicating emotion in music performance: A review and theoretical framework. In P. N. Juslin and J. A. Sloboda (Eds.), Music and emotion: Theory and research *(pp. 305-333). New York: Oxford University Press.*

Juslin, P. N., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in performance: The GERM model. Musicae Scientiae, Special Issue 2001-2002, *63-122.*

Schubert, E. (2003). Update of the Hevner adjective checklist. Perceptual and Motor Skills, 96, *1117-1122.*

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. Psyhology of Music, 19, *110-120.*

van Zijl, A. G. W., Toiviainen, P., & Luck, G. (2012). The sound of emotion: The effect of performers’ emotions on auditory performance characteristics. In E. Cambouropoulos et al. (Eds.), Proceedings of the ICMPC-ESCOM 2012, *Thessaloniki, Greece, 2012 (pp. 1064-1068).*

Zbikowski, L. M. (2010). Music, emotion, analysis. Music Analysis, Special Issue: Music and Emotion, 29, 1-3, *37-60.*

DIGITIZATION AND ANALYSIS OF SEASHORE’S HISTORICAL MUSIC PERFORMANCE SCORES

Johanna Devaney[1]
[1]*The Ohio State University, USA*

Time: 15:30

1. Background

In the first half of the 20th century, Carl Seashore and colleagues undertook extensive work in performance analysis. Many of the performance analyses were printed in their publications as performance scores, which display continuous fundamental frequency and loudness information as they evolve over time. The performance scores are rich in expressive performance parameters, but the researchers in Seashore’s time did not have the computational facilities or algorithms to perform more than simple analyses of them.

2. Aims: The first aim of this project is to use image-processing to accurately convert the visual data in Seashore’ performance scores into digital data points. The second is to analyze the timing, fundamental frequency, and loudness information encoded in the scores using both the original and modern analysis methods. The performance scores offer the opportunity to recover the data from the Seashore experiments for both validation of their initial calculations and more in-depth analyses.

3. Main contribution: This project is the first to apply optical plot reading to historic music performance scores. Currently, performance scores exist as illustrations of laborious work undertaken by the early pioneers of music performance research. By converting the performance scores to data, the wealth of performances examined by these scholars can be analyzed with contemporary approaches. This project focuses on the calculation of perceived pitch and tuning in four studies presented in the 1936 volume “Objective Analysis of Musical Performance”, which was edited by Carl Seashore. (1) Harold G. Seashore: An objective analysis of artistic singing, (2) Ray S.

Miller: The pitch of the attack in singing, (3) Arnold M. Small: An objective analysis of artistic violin performance, (4) Paul C. Greene: Violin performance with reference to tempered, natural, and Pythagorean intonation

4. Implications: The workflow created for this project will allow for the wealth of data available in performance scores not only by Seashore and colleagues but also other contemporary researchers, such as Milton Metfessel, to be revisited. The availability of this data in digital form would allow new research questions to be addressed, such as comparing historical performances for which recordings no longer exist to contemporary performances.

[4E] Music and Emotions

LOS ANGELES 14:00- 16:00 Tuesday 5 Aug 2014

REFRAMING EMOTIONS AND MUSIC

Finn Upham[1]
[1]*New York University, USA*
Time: 14:00

1. Background: A practical theory of music and emotion should be able to account for the fact that the same musical stimulus can be experienced many different ways by any given listener. Repeatedly, theories of emotion and music have restricted their focus on aspects of music listening are most obviously musical, particularly the signal, but many other factors contribute to listeners affective response.

2. Aims: Prepare a theoretical framework to explain listeners' emotional experiences of music which recognizes the many functions of music in daily life, from therapy to distraction to concert performance.

3. Main contributions: Factors determining a listener's experience of music fall into three categories: the social, the aesthetic, and the participatory. Music is, first and foremost, made by humans for other humans to engage with in the present moment. By hearing human action through music, our responses to it ranges as widely as our responses to each other, from indifference to passionate love or hate. Social factors related to emotion include performer intention (to calm, to woo), group association, and how these impact the listener. Aesthetic factors include appreciation of beauty and judgement of performers' actions in terms of effectiveness, quality, cleverness, and creativity. How we can and do act with music also effects how we feel about what we hear. Participatory factors include listener's understanding the music, what options they have for acting with the music, and whether they are allows to engage in being active with the music. Being active with can take many forms, from voluntary (or quasi-voluntary) amplification of empathy with the sentiments expressed in the music, to moving with the music, to making music with what is being performed.

4. Implications: A realistic theory of emotional responses to music must be extended to include scenarios outside of what is most convenient to measure, namely externally passive listening to unfamiliar music with little personal or social value. While a challenge to measure, it is imparative for empirical research into the unfolding experience of music to consider and quantify the active contributions of participants, given their musical history, mood, and inclination to engage at the moment of listening.

AFFECTIVE RESPONSES TO COMPLEXITY IN HARMONIC PROGRESSION BETWEEN MUSIC AND NONMUSIC MAJORS

Hyun Ju Chong[1], *Ga Eul Yoo*[1], *Soo Ji Kim*[1]
[1]*Ewha Womans University, Korea, Republic of*
Time: 14:30

1. Background: Musical elements, which determine the qualities of uncertainty, novelty, and complexity of music, have been considered to influence listeners' affective response to music. With regard to harmony, different chords in isolation and varied degree of interruption in expected harmonic progressions elicit different emotional responses. Also, an increasing number of studies support the findings that the intensity or sensitivity of these responses differ between musicians and nonmusicians. Still, different levels of chords and corresponding complexity of harmonic progression have been less extensively investigated as an agent that affects emotional responses, compared to the nature of isolated chords and the fulfillment of expected harmonic progressions.

2. Aims: This study aimed to examine listeners' responses to different levels of complexity in harmonic progression. Also, this study investigated whether those responses differ depending on the musical training of the listeners.

3. Method: Thirty seven music majors and 40 nonmusic majors participated in this study. The musical stimuli were composed of five different levels of harmonic progressions. After listening to each level of music, participants rated their responses on three 11-point semantic differential scales assessing the dimensions of activity, evaluation, and potency of emotion.

4. Results: Both music and nonmusic majors tended to perceive increased emotional activity, decreased evaluation, and increased emotional intensity in music with increased complexity of harmonic progression. However, increases in emotional activity with more complex harmonic progression occurred to a greater degree in music majors. They rated music with higher complexity more positively than did nonmusic majors. Also, perceived potency of emotion with increased complexity was considerably pronounced in the music major group.

5. Conclusions: The results of this study indicate similar traits in the responses of listeners from both groups; however, the music majors showed a greater degree of response than the nonmusic majors as the complexity in harmonic progression increased. This suggests that

music training and musical experience may prime listeners to more readily perceive increased energy, power, depth, and sensitivity of emotion with increased harmonic complexity.

NAMING THE ABSTRACT: BUILDING A REPERTOIRE-APPROPRIATE TAXONOMY OF AFFECTIVE EXPRESSION

Joshua Albrecht[1]
[1]*The University of Mary Hardin-Baylor, USA*
Time: 15:00

Studies measuring perceived musical affect have examined different groups of affects: "basic" emotions (Ekman 1992), arousal and valence terms (Schubert 1996), and eclectic lists of emotions (Eerola & Vuoskoski 2011). However, it is difficult to compare results between studies using different terms. Several recent projects have addressed this difficulty by building a general-purpose taxonomy of affective musical terms (Zentner, et al 2008, Schubert 2003). However, these studies assume that the same labels can be effectively used for many repertoires. Recently, Albrecht & Huron (2010) used a bottom-up approach to elicit piece-specific affective labels. This study seeks to expand this work by developing a taxonomy of terms to use for studying Beethoven's piano sonatas. The current report compares lists of affective terms gathered from three participant-generated approaches. In the first study, 28 participants each chose 20 terms deemed most relevant for "early Romantic piano music" from a list of 91 terms taken from three studies (Zentner, et al 2008, Schubert 2003, and Albrecht & Huron 2010). In the second study, 21 participants listened to 15 excerpts from Beethoven's piano sonatas and chose 5 terms for each excerpt from the same list. Finally, 43 participants provided free-response descriptions of the affect of the same excerpts, subjected to content analysis for derivation of affective terms. The results from these three studies were cross-referenced and a cluster analysis was used to determine a heuristic of affective categories. By cross-examining the results from the three different studies, nine affect terms are suggested for the Beethoven piano sonatas: angry, anticipating, anxious, calm, dreamy, exciting, happy, in love, and sad. The results from the three studies also illustrate how experimental paradigms influence participant-generated affective terms. The results from this study provide a taxonomy of affective terms that are repertoire-specific for the Beethoven piano sonatas. Allowing participants to provide free-response affect terms for this specific repertoire resulted in terms common to many studies (e.g. happy, sad), and also more unique terms that nonetheless seem relevant to this repertoire (dreamy, anxious, calm).

EMOTION RECOGNITION FROM MUSIC-INDUCED MOVEMENT

Edith Van Dyck[1]*, Pieter Vansteenkiste*[2]*, Matthieu Lenoir*[2]*, Micheline Lesaffre*[1]*, Marc Leman*[1]
[1]*Ghent University, Faculty of Arts and Philosophy, Department of Arts, Music and Theater Sciences, Institute for Psychoacoustics and Electronic Music, Belgium,* [2]*Ghent University, Faculty of Medicine and Health Sciences, Department of Movement and Sports Sciences, Motor Control and Learning, Belgium*
Time: 15:30

ABSTRACT

Emotions color all aspects of our interaction with music. Not only composing music or playing a musical instrument, but also perceiving sounds and responding to them implicates the involvement of human emotions. An interesting type of musical interaction, in particular with regard to emotion research, is dance, which is believed to facilitate the expression of several different emotions in a non-verbal way. In this study, the aim was to examine emotion perception from dance movement. Thirty participants observed a selection of silent videos showing depersonalized avatars of dancers moving to an emotionally neutral musical stimulus after emotions of either sadness or happiness had been induced. After every film clip, the participants were asked to assess the emotional state of the dancer. Results revealed that the emotional state of the dancers was successfully identified. In addition, emotions were more often recognized for female dancers than for their male counterparts. Finally, results of eye tracking measurements showed that observers primarily focused on movements of the trunk when decoding emotional information from dance. The findings of this study show that induced emotions can be successfully recognized from dance movement. They also illustrate the significance of emotions in the coupling between music perception, cognition, and action.

1. INTRODUCTION

One of the main reasons why music is so abundantly present in almost every possible activity of our lives can be traced back to its emotional effects (Juslin & Laukka, 2004). Since music is often considered as a language of emotions (Cooke, 1959), it should not come as a surprise that a multiplicity of studies have been devoted to the exploration of the link between music and emotion (Juslin & Sloboda, 2001). Since it has been shown that cognition, perception, and action are tightly coupled (Leman, 2007), it is self-evident that when we interact with music, corporeal articulations are involved. Moreover, neuroscientific research has demonstrated that brain regions responsible for motor activities, as well as emotional responses, are activated through musical interaction (Blood & Zatorre, 2001; Janata & Grafton, 2003; Koelsch, 2010; Peretz & Zatorre, 2005). Hence, a disembodied, Cartesian mind as such does not exist. Therefore, either when dealing with emotion effects that are rooted in musical interaction or with effects of sound on affective experience, the impact of the body can simply not be ignored.

An interesting example of embodied interaction with music, in particular with regard to emotion research, is music-induced movement, or dance, which is believed to facilitate the expression of several different emotions in a non-verbal way (Levy, 1988). Moreover, dance is one of the oldest forms of cultural expressions (Sachs, 1937). The ability to successfully decode emotions from dance movement is already present from early childhood onwards. In addition, children as young as five years old have demonstrated to be capable of decoding the intensity of the emotions expressed by dancers when observing videos of adults moving expressively to music

(Boone & Cunningham, 1998). Previous research concerning emotion recognition from dance commonly used videos of actors or dancers portraying emotions such as happiness, sadness, fear, disgust and anger (Boone & Cunningham, 1998; Camurri, Lagerlöf, & Volpe, 2003; de Meijer, 1989; Lagerlöf & Djerf, 2000; Montepare et al., 1999; Pollick et al., 2001). The assumption underlying the selection of emotions presented through acting in previous studies is that actors are typically believed to be experts in displaying emotional information corporeally and that acted emotions parallel emotions experienced in real life (Gross, Crane, & Fredrickson, 2010). However, one might argue that these acted actions are exaggerated and should rather be regarded as symbolic portrayals of the emotions at issue. In addition, a number of studies have revealed that not all actors generate equally identifiable, emotionally expressive dance movements (Gross, Crane, & Fredrickson, 2010; Montepare et al., 1999; Wallbott, 1998). Therefore, instead of considering emotions portrayed through acting, the current study applies emotion induction techniques in order to examine the recognition of emotions from dance movement, as induced emotions are supposedly equivalent to naturally occurring emotions (Jallais & Gilet, 2010). Furthermore, contrary to previous studies, which have made use of choreography to study expressive dance movements (Brownlow et al., 1997; Camurri, Lagerlöf, & Volpe, 2003; Montepare et al., 1999; Wallbott, 1998), the dancers in this study were able to move freely as we believe that this facilitates more accurate expression of emotion.

In a previous study by Van Dyck et al. (2013), which examined the impact of induced emotions (happiness and sadness) on dance movements of participants dancing freely to emotionally neutral music by measuring and analyzing the kinematics of the movements, evidence of the effect of emotion induction on dance movement was provided. In the current study, it is investigated whether human observers are able to decode induced emotions from dance. Since both adults and children have been proved to master the ability to decode portrayed emotions from full body movements (Boone & Cunningham, 1998; Dittrich, 1996; Lagerlöf & Djerf, 2000; Shikanai, Sawada, & Ishii, 2013), we expect that observers are also capable of distinguishing between induced emotions manifested through dance movement. In addition, since little is known about visual search patterns people use to gather information on which emotion recognition is built, the behavioral method of eye tracking is used to examine the direction of the observers' focus. We also presume that female participants perform better in the recognition task than their male counterparts, as women are believed to be superior in understanding others' emotions (Hampson, van Anders, & Mullin, 2006). In addition, as women generally experience and express emotions more intensely than do men (Donges, Kersting, & Suslow, 2012), we expect the participants to have a higher success rate when judging the emotional state of female than of male dancers. Finally, as Van Dyck et al. (2013) revealed that the differences between the emotion conditions were mainly detectable from hand movement, we expect observers to primarily focus on the gestures of the dancers' hands.

2. METHOD

2.1. Participants

A total of 30 adult observers (15 females, 15 males) took part in the study. The average age of the observers was 27.23 years (*SD* = 3.43). 60 % had received musical training, and of those, the average time spent in musical training was 6.23 years (*SD* = 6.70). About half of the observers (56.70%) had also trained in dance and spent, on average, 2.67 years (*SD* = 4.02) in dance training. The observers received no compensation for participating in the study.

2.2. Materials and stimuli

Stimuli. Sixteen video clips were used in this study. The video clips were recorded with motion capture cameras in a previous study (Van Dyck et al., 2013), in which the effect of induced emotions on the kinematics of dance movement was studied. In that particular study, non-expert dancers were induced to feel emotional states of either happiness or sadness and then danced intuitively to an emotionally neutral piece of music. Each participant performed both conditions. For the current study, silent video clips showing depersonalized, androgynous avatars were created from the movement data of the dancers. Each of the 16 video clips, consisted of a pair of dance performances (presented side-by-side and played simultaneously); one of a dancer in the happy condition and one of the same dancer in the sad condition. Both the order of the video clips and of the emotion conditions was randomized. Half of the dancers were female and all emotional states were presented an equal number of times on the left and on the right side of the screen. Each video clip had a duration of 10 seconds. Three practice clips preceded the actual clips that had to be rated. The video clips were presented on a 22-inch computer monitor.

Eye tracking. Eye movements were recorded using a Remote Eye tracking Device (RED) by SensoMotoric Instruments (SMI). To calculate the time participants watched the different regions of the body, dynamic Areas Of Interest (AOIs) were coded on the video clips. Once the AOIs were coded, dwell-time percentages (percentage of time the eyes were directed towards the AOI) were retrieved. In order to control for AOI size, dwell-time percentages were normalized for the size of the specific body area. All clips were coded with following AOIs: chest, feet, head, hips, arms, legs, and hands.

2.3. Procedure

Participants were asked to watch the video clips and to fill out a short questionnaire after each video. In the questionnaire, they were asked to point out which of the two dance performances was happy and which one was sad. At the end of the experiment, the observers were asked to fill out a second questionnaire, which contained questions concerning their music and dance background.

3. RESULTS

3.1. Emotion recognition

A one-sample chi-square test revealed that observers were able to recognize the correct emotion from the dance movements, $\chi^2(1) = 255.21$, $p < .001$. Based on the odds ratio, the odds of observers recognizing the correct emotion were 6.38 times higher than making the wrong decision concerning the emotional state of the dancers. In addition, the effect of the order of the videos, the order of the emotions, the sex of the dancers, and the sex of the observers was examined. Pearson's chi-square tests revealed no significant associations between the order of the videos, $\chi^2(2) = .64$, $p = .89$, the order of the emotions, $\chi^2(1) = 1.78$, $p = .18$, or the sex of the observers, $\chi^2(1) = 1.14$, $p = .29$. However, there was a significant association between the sex of the dancers and the results of the emotion recognition task, $\chi^2(1) = 28.47$, $p < .001$. Based on the odds ratio, the odds of recognizing the correct emotional state were 5.38 times higher when observers were watching female dancers than when they were observing male dancers.

3.2. Eye-tracking

Validation tests showed an average accuracy of 0.37° (*SD* = 0.12) and the average tracking ratio (% of time eye movements was actually measured) was 89.90 % (*SD* = 3.70), which signifies that the tracking data of all participants proved to be accurate enough in order to be analyzed statistically.

AOIs were analyzed in order to check the direction of the focus of the observers' gaze. A Kolmogorov-Smirnov test showed that the assumption of normality could not be accepted. A Friedman's ANOVA with the AOIs as test variables showed a significant difference in mean dwell time between the different AOIs, $\chi^2(6) = 128.69$, $p < .01$. Wilcoxon tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a .00048 level of significance. Mean dwell time was significantly higher for the chest than for the head, $Z(29) = -4.76$, hips, $Z(29) = -4.78$, arms, $Z(29) = -4.78$, hands, $Z(29) = -4.78$, legs, $Z(29) = -4.78$, and feet, $Z(29) = -4.78$. In addition, mean dwell time was significantly higher for the arms than for the hips, $Z(29) = -4.68$, hands, $Z(29) = -4.78$, legs, $Z(29) = -4.39$, and feet, $Z(29) = -4.78$. Finally, mean dwell time was significantly lower for the feet than for the head, $Z(29) = -4.33$, chest, $Z(29) = -4.78$, hips, $Z(29) = -4.61$, arms, $Z(29) = -4.78$, hands, $Z(29) = -4.70$, and legs, $Z(29) = -4.78$. An overview of the mean dwell time for the different AOIs is presented in Figure 1.

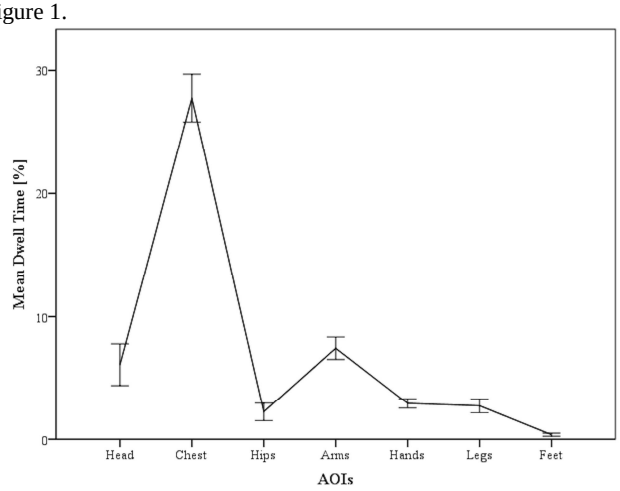


Figure 1: Mean dwell time for the AOIs. Data presented are mean ± *SE*.

4. DISCUSSION

In this study, it was examined whether human observers are able to decode emotional content from corporeal articulations of dancers moving to an emotionally neutral piece of music after emotions (happiness or sadness) had been induced. The results revealed that the participants were indeed capable of identifying the intended emotion from the dance movements. This finding is in accordance with results of previous studies on emotion recognition from dance, examining portrayed emotions. For instance, Boone and Cunningham (1998), Montepare et al. (1999), and Shikanai et al. (2013) showed that portrayed emotions expressed in dance could be accurately identified by observers. Therefore, our results resonate with findings of previous studies, but also extend them by showing that, in addition to portrayed, or 'acted', emotions, also induced emotions can be decoded from dance.

As it is well documented that women are better at understanding and considering the feelings and needs of others compared to men (Hampson, van Anders, & Mullin, 2006), we presumed female observers to recognize the induced emotions more often than their male counterparts. However, our results did not support this premise. On the other hand, a significant association with recognition accuracy was unveiled regarding the sex of the dancers, as the emotional state was more often recognized for female dancers than when male dancers were being observed. This suggests that women are more proficient in expressing their personal feelings in a corporeal manner compared to men. This is in accordance with a fairly substantial body of research, which has demonstrated that women are more emotional, and both experience and express emotions more intensely than men (Donges, Kersting, & Suslow, 2012).

As the hands are generally believed to have a privileged role in music-related gestures (Godøy, 2010), and since Van Dyck et al. (2013) obtained more significant differences for the hands than for any other body part, we expected that participants would mainly focus on hand gestures. However, eye tracking data from the current study unveiled a specific focus of the observers on the chest area. Although this finding did not fit our expectations, several explanations arise from the data. First, most of our body movements tend to start from the more proximal segments and develop towards the more distal limbs (Chapman, 2008). This implies that information on changes in movement direction or acceleration might be readily seen in the trunk/shoulder area first. Second, even though the observers' general focus is on the chest area, gestural information concerning other body parts is not necessarily disregarded as the participants are still capable of perceiving movements of other parts of the body, in relation to the chest. In this specific experimental set-up, most of the movements of the arms and legs were still within the useful field of view when focusing on the chest. The reported gaze behavior therefore suggests that dancers' emotions were analyzed using an extended visual span and parafoveal processing. This visual strategy has been described in fields such as sports (Gegenfurtner, Lehtinen, & Säljo, 2011) and radiology (Kundel et al., 2007) and has been labeled as 'the holistic model of image perception'. This visual strategy suggests that observers do not pay attention to the head, the hands, the hips, etc. as separate parts of the body, but rather see the human body as one entity. Our finding with regard to the focus on the chest accords with previous research on emotions and body posture as several studies have emphasized the importance of the posture and position of the torso as indicators for emotional content. Schouwstra and Hoogstraten (1995), for instance, used stick drawings of armless figures and varied the positions of the spine (and head). Their study revealed that upright postures were judged more positively, and forward-leaning postures more negatively.

Previous research revealed that, even though emotion recognition is not fully matured until early teenage years (Tonks et al., 2007), children as young as about five years old perform above chance in successfully perceiving emotional information from body language. Moreover, they are capable of decoding the intensity of the emotions expressed through dance movement when observing videos of people moving to music (Boone & Cunningham, 1998). As it is believed that the duration of negative emotional states decreases with age (Larcom & Isaacowitz, 2009), the current study only considered a specific age group (from 24 to 34 years of age). However, a future study could investigate whether also children or observers in other age groups are capable of successfully recognizing induced emotions from free dance movements.

With regard to the stimuli used in this study, each film clip consisted of two dance performances executed by the same individual. Pairs of performances of one and the same dancer were employed in order to ensure optimal control over possible confounding effects caused by characteristics of the dancers, such as the their personality, gender, and proclivity to dance, since these features have been proved to influence human dance movement significantly (Passmore & French, 2001; Risner, 2009). However, the skill to discriminate between happy and sad performances of the same dancer might be different from the ability to differentiate between happy and sad dances across different performers. Yet, this is a matter of some speculation and would benefit from further study.

In summary, this experimental study examined whether observers are able to decode induced emotions from dance movement. Our results are in tune with results of similar studies, but they also extend previous research, showing that, in addition to portrayed emotions, also induced emotions can be perceived from unchoreographed dances by adult observers. Moreover, this study shows that female dancers are better at communicating emotional meaning corporeally than their male counterparts. Finally, the results of this study unveiled that observers generally focus on movements of the chest when decoding emotional information from dance movement.

5. REFERENCES

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. PNAS, 98, *11818–11823.*

Boone, R. T., & Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: the development of cue attunement. Developmental Psychology, 34, *1007–1016.*

Brownlow, S., Dixon, A. R., Egbert, C. A., & Radcliffe, R. D. (1997). Perception of movement and dancer characteristics from point-light displays of dance. The Psychological Record, 47, *411–421.*

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. The International Journal of Human-Computer Studies, 59, *213–225.*

Chapman, A. (2008). Biomechanical analysis of fundamental human movements. *Champaign, IL: Human Kinetics.*

Cooke, D. (1959). The language of music. London: Oxford University Press.

Darwin, C. (1872). The expression of the emotions in man and animals*. London, UK: John Murray.*

de Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. Journal of Nonverbal Behavior, 13, *247–268.*

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. Perception, 25, *727–738.*

Donges, U.-S., Kersting, A., & Suslow, T. (2012). Women's greater ability to perceive happy facial emotion automatically: Gender differences in affective priming. PLoS ONE, 7, *e41745.*

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. Psychological Bulletin, 94, *100–131.*

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In Cole, J. (Ed.), Nebraska symposium on motivation. *Lincoln, NE: University of Nebraska Press.*

Gegenfurtner, A., Lehtinen, E., & Säljo, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. Educational Psychology Review, 23, *523–552.*

Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. Speech Communication, 40, *189–212.*

Godøy, R. (2010). Gestural affordances of musical sound. In Godøy, R., & Leman, M. (Eds.), Musical gestures: Sound, movement, and meaning. *New York, NY: Routledge.*

Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2010). Methodology for assessing bodily expression of emotion. Journal of Nonverbal Behavior, 34, *223–248.*

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. Evolution and Human Behavior, 27, *401–416.*

Jallais, C., & Gilet, A.-L. (2010). Inducing changes in arousal and valence: Comparison of two mood induction procedures. Behavior Research Methods, 42, *318–325.*

Janata, P., & Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. Nature Neuroscience, 6, *682-687.*

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. Journal of New Music Research, 33, *217–238.*

Juslin, P. N., & Sloboda, J. A. (Eds). (2001). Music and emotion: Theory and research. *Oxford: Oxford University Press.*

Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. Trends in Cognitive Sciences, 14, 131–137.

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. Radiology, 242, *396–402.*

Lagerlöf, I., & Djerf, M. (2000). Communicating emotions: Expressiveness in modern dance. International Journal of Psychology, 35, *225–225.*

Larcom, M. J., & Isaacowitz, D. M. (2009). Rapid emotion regulation after mood induction: Age and individual differences. The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 64, *733–741.*

Laukkanen, A.-M., Vilkman, E., Alku, P., & Oksanen, H. (1997). On the perception of emotions in speech: The role of voice quality. Logopedics Phoniatrics Vocology, 22, *157–168.*

Leman, M. (2007). Embodied music cognition and mediation technology. *Cambridge: MIT Press.*

Levy, F. J. (1988). Dance movement therapy: A healing art. *Reston, VA: The American Alliance for Health, Physical Education, Recreation and Dance.*

Likowski, K. U., Weyers, P., Seibt, B., Stöhr, C., Pauli, P., & Mühlberger, A. (2011). Sad and lonely? Sad mood suppresses facial mimicry. Journal of Nonverbal Behavior, 35, *101–117.*

Luck, G., Saarikallio, S., Burger, B., Thompson, M. R., & Toiviainen, P. (2010). Effects of the Big Five and musical genre on music-induced movement. Journal of Research in Personality, 44, *714–720.*

Lundqvist, L.-O., & Dimberg, U. (1995). Facial expressions are contagious. Journal of Psychophysiology, 9, *203–211.*

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. Psychological Bulletin, 126, *424–453.*

Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. Journal of Nonverbal Behavior, 23, *133–52.*

Niedenthal, P. M. (2007). Embodying emotion. Science, 316, *1002–1005.*

Passmore, A. E., & French, D. (2001). Development and administration of a measure to assess adolescents' leisure participation. Adolescence, 36, *67–75.*

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. Annual Review of Psychology, 56, 89–114.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. Cognition, 82, *B51–B61.*

Risner, D. (2009). What we know about boys who dance: The limitations of contemporary masculinity & dance education. In Shay, A., & Fisher, J. (Eds.), When men dance: Choreographing masculinities across borders. *New York, NY: Oxford University Press.*

Sachs, C. (1937). World history of the dance. *New York, NY: Norton.*

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, 40, *227–256.*

Schouwstra, S. J., & Hoogstraten, J. (1995). Head position and spinal position as determinants of perceived emotional state. Perceptual & Motor Skills, 81, *673–674.*

Shikanai, N., Sawada, M., & Ishii, M. (2013). Development of the movements impressions emotions model: Evaluation of movements and impressions related to the perception of emotions in dance. Journal of Nonverbal Behavior, 37, *107–121.*

Sroufe, L. A. (1996). Emotional development. *New York, NY: Cambridge University Press.*

Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007). Assessing emotion recognition in 9-15-years olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. Brain Injury, 21, *623–629.*

Van Dyck, E., Maes, P.-J., Hargreaves, J., Lesaffre, M., & Leman, M. (2013). Expressing induced emotions through free dance movement. Journal of Nonverbal Behavior, 37, *175–190.*

Wallbott, H. G. (1998). Bodily expression of emotion. European Journal of Social Psychology, 28, *879–896.*

Session 5

[5A] Tonality vs Atonality

SEOUL 16:30- 18:00 Tuesday 5 2014

THE LIMITS OF TONALITY: SEMITONAL MODULATIONS IN CHOPIN'S MUSIC

Heewon Chung[1]
[1]*University of Michigan, USA*
Time: 16:30

This paper deals with theoretical problems raised by the coordination of voice leading with modulation to remote tonal regions in the music of Chopin. Compositional techniques of bold modulation in nineteenth-century music frequently inspire discussion about the boundaries of tonality, since chromatic anomalies motivate analysts to incorporate new perspective. Many scholars have noticed Chopin's harmonic boldness and his extensive use of third relationships, but have not fully explored his use of semitonal relationships. I claim that semitonal relationships play an important role in Chopin's oeuvre. In particular, I focus on those modulations that involve a transformation between scale-degree 1 and 7, analyzing Chopin's Nocturne in Bb Minor, Op. 9, No. 1, Nocturne in C# Minor, Op. 27, No. 2, and Ballade Op. 52. I call this "leading-tone modulation," since it occurs when the tonic in one key changes into the leading tone of the new key or vice versa. The approach taken here will differ somewhat from that of scholars who theorize semitonal modulation purely in terms of chromatic space, since an overarching diatonic space will continue to operate. Rather, this paper explain the tonal irregularities in this kind of modulation as involving a voice-leading gap arising from the linear use of chromatic substitution and unusual counterpoint; perceiving these may require retrospective hearing. I believe that semitonal relationships provide analysts a powerful vantage point from which to examine Chopin's style. Even though Chopin is not the only one to use semitonal modulations, the way he uses them is distinctive, as analysis can show. My perspective also hopes to address the listener's tonal disruption that the semitonal modulation often arouses since it is usually associated with musical event involving unusual voice-leading. This has cognitive implication since unusual voice leading will often disorient a passage's tonal meaning. Semitonal relations engage significant current issues in tonal theory, such as diatonic-chromatic interaction, enharmonicism, and phenomenological hearing. I believe these theories of tonal relationship can be extend and enriched through a close consideration of a composer's individual style and the particular analytical context within which semitonal relations arise.

IMPLICIT LEARNING OF MUSICAL STRUCTURE IN TWELVE-TONE MUSIC: WHAT DO NOVICE LISTENERS GLEAN FROM THE MUSICAL SURFACE?

sempre: Society for Education, Music and Psychology Research

Jenine L. Brown
Madonna University

Time: 17:00

ABSTRACT

Many have described twelve-tone music as difficult to aurally comprehend (e.g. Meyer 1967, Huron 2006). This study addresses such claims by investigating what novice listeners can implicitly learn when hearing a recording of a twelve-tone composition. Krumhansl (1990) has argued that listeners unfamiliar with a musical style attune to the distribution of pitch occurrences, with the most frequent pitch providing a reference point. However, in Anton Webern's Concerto, Op.24/iii, each pitch occurs nearly the same number of times. Because the distribution of pitches in this twelve-tone work is flat, this study investigates whether listeners will perceive its recurring intervals instead. After passive exposure to the composition, musician listeners (n=12) with no formal training in non-tonal music theory demonstrated implicit learning of the frequent intervals (and pairs of intervals) in both forced-choice and ratings tasks. This interval-based view of twelve-tone perception informs an analytical approach also illustrated in this study.

1. INTRODUCTION

Carol Krumhansl writes that when music "organizes the materials around a few reference pitches, this organization is readily apprehended by listeners with little previous experience with the musical style" (1990, 270). Indeed, Kessler, Hansen, & Shepard (1984), Castellano, Bharucha, & Krumhansl (1984), and Krumhansl, Sandell, & Sergeant (1987) have suggested that listeners unfamiliar with a musical style attune to the distribution of pitch.

In twelve-tone music, however, each pitch often occurs the same number of times as the other. For example, in Anton Webern's twelve-tone composition titled *Concerto for Nine Instruments*, Op. 24, third movement, each pitch is heard between 8 and 9% of the time (Figure 1).

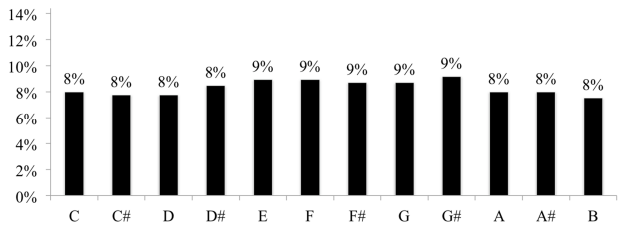


Figure 1: Percentage that each pitch occurs in Webern's Concerto, Op. 24, third movement.

This distribution of pitch in twelve-tone music contrasts with that of tonal music, in which scale degrees ^1, 3, and 5 (and other diatonic tones) often occur more frequently than other chromatic pitches. For example, Figure 2 shows the pitch distribution of compositions transposed to the key of C major, compiled by Youngblood (1958) and Knopoff & Hutchinson (1983) from works such as Schubert Lieder and Mozart arias.

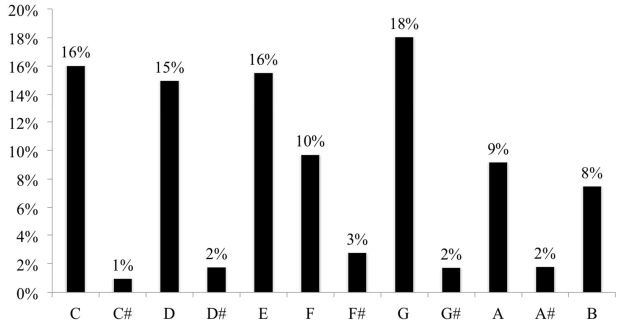


Figure 2: Percentage that each pitch occurs in works in the key of C major (adapted from data presented in Krumhansl, 1990, 67).

Because the distribution of pitch in twelve-tone works is often flat (and because this distribution differs from that presented in tonal music), scholars have questioned the listener's ability to aurally comprehend twelve-tone music (Meyer, 1967; Lerdahl & Jackendoff, 1983; Huron, 2006). Nevertheless, numerous studies have investigated the perception of non-tonal music, including testing the listener's ability to perceive the similarity of atonal chords (Bruner, 1984; Samplaski, 2004; Kuusi, 2003), to identify and learn row forms and relationships (Krumhansl, Sandell, & Sergeant, 1987; Dienes & Longuet-Higgins, 2004), to recognize avoided pitch repetition in twelve-tone music (Krumhansl, Sandell, & Sergeant, 1987; Ockelford & Sergeant, 2013), and to learn surface characteristics of Messiaen's serial music (Krumhansl, 1991). These experiments have almost exclusively studied the perception of pitch.

2. INTERVALS IN 12-TONE MUSIC

The present study instead explores the listener's ability to learn intervals in twelve-tone music. Indeed, while there is infrequent pitch repetition in twelve-tone compositions, some compositions contain repetitions of elements in other musical domains. For example, pitch intervals 4, 8, 11, and 13 predominate the musical surface of Webern's Op. 24.10 Furthermore, melodic trichords are common on the musical surface of Op. 24/iii, and most of these three-note melodies contain 11 or 13 semitones between one adjacent pair of notes, and 4 or 8 semitones between the other pair. These intervals can be called "frozen," as a nod to the research that has described Webern's use of frozen pitches (e.g. Westergaard, 1963; Bailey, 1983; Alegant, 2001). That is, in works with frozen pitches, a pitch-class is heard as just one pitch on the musical surface: in Webern's Op. 27/ii, pitch-class A is always and only heard as A4. In Webern's Op. 24/iii, however, certain interval-classes are presented as one (or two) pitch interval(s) on the musical surface. For example, interval-class 1 is exclusively heard throughout the movement as 11 or 13 semitones.

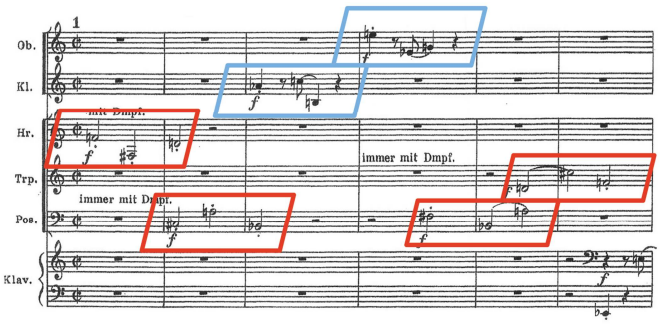


Figure 3: Motives and near-motives in Webern's Op 24/iii (mm. 1-6).

Figure 3 illustrates these frozen intervals as they appear on the musical surface of Op. 24/iii. The opening six measures contain trichords, all of which are members of SC[014]. The trichords boxed in red contain intervals 8 and 11. Because this interval pair occurs most often in the movement (occurring in 31 of the 44 melodic trichords), it can be called the motive. The motive in this movement, then, is defined as a three-note melody that can start on any note and contains intervals 8 and 11 (in any order), where one interval ascends and the other descends. The remaining trichords on the musical surface of the movement are also members of SC[014], like the motive, but these trichords (boxed in blue) are expressed with different intervals and occur less often than the motive. These intervallic variants can be called "near-motives."

10 Throughout this paper, the distance between two pitches is described in semitones (also called pitch intervals).

Such an interval-based approach to this music is not novel, as is clear when Headlam writes, "the interpretation of non-tonal music in terms of intervals rather than pitch-class content is basic to our understanding of this music" (2012, 136). However, the present study is unique in its experimental test of the perception of intervals in twelve-tone music.

3. EXPERIMENT

The following music listening experiment tested the listener's ability to learn recurring intervals in Webern's Op. 24, third movement.

3.1. Method

Participants. Participants (n=12) were freshmen and sophomore music majors at the Eastman School of Music, and had no formal training in non-tonal music. Participants (6 F, 6 M) had a mean age of 19.9 years (SE=0.34) and were preparing for professional music careers. They reported to have studied their instruments for an average of 12.9 years (SE=1.22). Participants were volunteers, solicited by class announcements in freshmen and sophomore music theory and aural skills classes, who were paid for their participation. At the conclusion of the experiment, all participants were given a 24-item pitch-labeling test. Those who scored at least 93% correct were deemed to have absolute pitch, permitting half-step errors. Of the 12 participants, three had absolute pitch.

Apparatus. The experiment took place in a small room. Participants were tested individually, and responded to prompts on an iMac computer while hearing the musical stimuli through high-quality headphones. Participants were allowed to adjust the loudness of the stimuli during the experiment. Subject responses were collected using Experiment Creator (version 1.4), a program designed by William F. Thompson (http://www.psy.mq.edu.au/me2/index.php/site/creator/).

Stimuli. Stimuli presented during the test phases were generated by the Sibelius notation software program and exported as .wav files. All stimuli were synthesized piano tones, and each note was heard individually for 0.33 seconds. Musical stimuli ranged from C3 and G5, but primarily spanned A3 to A4.

Procedure. An important question is whether listeners brought to the experiment any biases for or against particular stimuli–biases gleaned from implicit learning in everyday music listening. This was tested in the pre-exposure tests described below. The entire experiment contained seven different phases:

Table 1: The 34 probe-melodies rated before and after familiarization.

Probe-melodies rated:	Intervals (in semitones)
Motives (6 trials). Heard 67% of all melodic trichords in Op. 24/iii.	+8-11, +11-8, -8+11 (2x), -11+8 (2x)
Near-motives (6 trials). Heard 20% of all melodic trichords in Op. 24/iii.	-13+4, +4-13, +8+13, +13+8, +11+4, -11-4
Large dissonant non-motives (4 trials). Heard 0% in Op. 24/iii.	+1-14, +6+8, -15+9, -4+20
Large consonant non-motives (6 trials). Heard 0% in Op. 24/iii.	-9+7, +15-7, -14-2, +7+7, -1+15, +16+1
Small consonant non-motives (4 trials). Heard 0% in Op. 24/iii.	-7+3, -2+5, -2-2, +2+5
Intervals heard often in Op. 24/iii (4 trials).	+11, -11, +13, -13
Intervals heard infrequently in Op. 24/iii (4 trials).	+10, -10, +14, -14

1. PRE-EXPOSURE FORCED-CHOICE TEST: In this initial phase of the experiment, listeners heard two melodic trichords and were asked to select the one that occurs more often in music heard throughout their lifetimes. There were 2 seconds of silence between each trichord, and listeners could take as long as they wanted to make their selection. Listeners heard 28 forced-choice trials, randomized differently for each participant.
2. PRE-EXPOSURE PROBE-MELODY TEST: Then, listeners rated 8 probe-intervals and 26 probe-trichords on how often they occur in music heard throughout their lifetimes (on a scale from 1 to 7, where 7 is most often). These trials sounded like two or three-note melodies, played in a synthesized piano timbre, where each note was 0.33 seconds and there were no pauses between notes. Trials were randomly ordered for each participant, and the probe-melodies had varying contours, starting pitches, and intervals. All but one of the 12 trichordal set-classes were rated. The 34 probe-melodies rated are listed in Table 1; trichords are shown as two intervals.
3. FIRST FAMILIARIZATION PHASE: Participants then heard a recording of Webern's Op. 24/iii (Webern, 1991), repeated seven times to last for 9'33". A few seconds of silence were cut from the original track so that there was no pause between repetitions. Participants were not told what to listen for, or anything about the composition they were hearing.
4. POST-EXPOSURE FORCED-CHOICE TEST: Listeners were then asked to discriminate between a trichord heard often in Webern's Op. 24/iii versus a trichord that occurred less often (or never) in a series of 28 two-alternative, forced-choice trials. These trials were identical to the trials from the pre-exposure forced-choice test, except this time listeners were asked to select the melody that occurred more often in music they heard during the familiarization phase (rather than which one occurred more often in music heard throughout their lifetimes). Trials were randomly ordered for each participant, and stimuli was chosen from the probe-melodies listed in Table 1. Listeners discriminated between trichords in five different types of trials. In the following four conditions, it was

hypothesized that listeners would choose the motive, because it occurred more often during familiarization: motives vs. large dissonant non-motives (4 trials), motives vs. large consonant non-motives (6 trials), motives vs. small consonant non-motives (4 trials), and motives vs. near-motives (4 trials). Large dissonant non-motives, large consonant non-motives, and small consonant non-motives did not occur during familiarization, but near-motives did occur (albeit less often than the motive). The final condition contained comparisons that were foils, inserted randomly so that listeners were less likely to notice that the motive always appeared as one of the two forced-choice options in the above conditions. In these foils, listeners compared near-motives vs. large consonant non-motives (4 trials) and near-motives vs. large dissonant non-motives (6 trials).

5. SECOND FAMILIARIZATION PHASE: Listeners heard the 9'33" familiarization phase again, which was identical to the first familiarization phase.
6. POST-EXPOSURE PROBE-MELODY TEST: In this test, listeners rated the same 34 trials from the pre-exposure probe-melody test. This time, the prompt asked listeners to rate how often the short melody occurred in the music heard during familiarization. Trials were randomized differently for each participant. Motives, near-motives, and intervals 11 and 13 occurred within Webern's composition, but all other short melodies (i.e. large dissonant non-motives, large consonant non-motives, small consonant non-motives, and intervals 10 and 14) never occurred. It was hypothesized that listeners would rate motives higher than all other trichords rated, because motives were heard most often during Op. 24/iii, occurring 67% of all melodic trichords heard in the composition. It was similarly hypothesized that intervals +/-11 and 13, which were heard prominently throughout the composition (occurring within every group of three notes in the composition), would be rated higher than similar dissonant and large intervals not heard in Webern's Op. 24/iii, such as intervals +/-10 and 14.
7. POST-TEST QUESTIONNAIRE: At the end of the experiment, listeners were administered a short test for absolute pitch and a questionnaire about their background and musical training.

3.2. Results

Data for the three AP possessors were compared against the non-AP participants, and no main effect for AP was found. Therefore, all data were combined for analysis.

The primary research question was whether listeners would implicitly learn the common intervals within Webern's Op. 24/iii. In the case of the probe-intervals heard in the post-exposure probe-melody test, listeners rated intervals 11 and 13 (intervals that occurred often during familiarization) at M=5.15 (SE=0.27). These intervals were rated significantly higher than ratings for intervals 10 and 14, which were rated at M=4.33 (SE=0.23), $t(11)=2.85$, $p<0.05$, paired, two-tailed. Difference scores were also used to analyze listener responses, in an attempt to remove the bias for intervals obtained prior to experimentation. The d-score was equal to the listener's rating of a probe-melody after hearing the familiarization phases, minus their rating of that same melody taken before hearing the familiarization phases. The d-score for probe-intervals +/-11 and 13 was 0.28 (SE=0.21), meaning that listeners rated these melodic intervals as occurring slightly more often in Op. 24/iii than in music heard throughout their lifetimes. This d-score was significantly higher than the d-score of probe-intervals +/-10 and 14, which was -0.29 (SE=0.23), $t(11)=3.06$, $p<0.05$, paired, two-tailed. These results suggest that listeners implicitly learned the often-occurring intervals 11 and 13 from familiarization, despite any biases they had for these intervals prior to experimentation.

Responses for the probe-trichords were also analyzed (Figure 4). In the post-exposure probe-melody test, listeners rated motives (M=5.38, SE=0.16) significantly higher than all other trichords rated, suggesting implicit learning of the most common interval pairing in Webern's Op. 24/iii.11

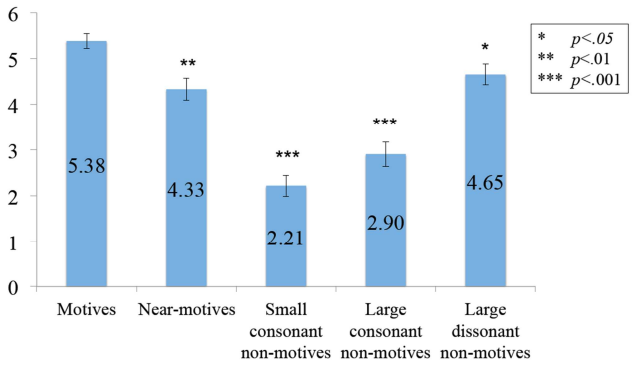


Figure 4: Average ratings of trichords from the post-exposure probe-melody test.

11 Listener ratings of motives were significantly higher than ratings of small consonant non-motives (M=2.21, SE=0.23, $t(11)=9.39$, $p<0.001$), large consonant non-motives (M=2.90, SE=0.27, $t(11)=7.04$, $p<0.001$), large dissonant non-motives (M=4.65, SE=0.23, $t(11)=2.67$, $p<0.05$), and near-motives (M=4.33, SE=0.24, $t(11)=3.35$, $p<0.01$), all determined by paired, two-tailed t-tests.

Analysis of d-scores for trichords rated before and after familiarization indicates that listeners learned the motives despite any biases they had for these melodies prior to experimentation. D-scores for motives (M=2.08, SE=0.25) were significantly higher than d-scores for small consonant non-motives (M= -2.14, SE=0.32), large consonant non-motives (M= -2.20, SE=0.39), large dissonant non-motives (M=1.33, SE=0.30), and near-motives (M=1.07, SE=0.31).12

Furthermore, listeners learned the distribution of interval pairs heard during familiarization, as average listener ratings of trichords from the post-exposure probe-melody test correlated with how often each interval pair occurred during Op. 24/iii at $r(24)=0.716$, $p<0.001$. Difference score means also correlated with these frequencies of occurrence at $r(24)=0.618$, $p<0.001$.

In the case of the post-exposure forced-choice test, listeners demonstrated that they learned the motive from familiarization, correctly discriminating between it and a trichord that occurred less often (or not at all) during Webern's Op. 24/iii. For example, listeners chose motives over near-motives (which also occurred in Webern's composition, albeit less often than the motive) at M=0.71 (SE=0.06); this mean differed significantly from chance (0.50), $t(11)=3.46$, $p<0.01$ (two-tailed, paired samples). Listeners also discriminated between motives and non-motives (which never occurred during familiarization) at above-chance levels: listeners chose motives over small consonant non-motives at M=0.96 (SE=0.03), significantly above-chance ($t(11)=16.32$, $p<0.001$), and listeners chose motives over large consonant non-motives at M=0.82 (SE=0.05), significantly above chance ($t(11)=6.17$, $p<0.001$). Listeners also selected motives over large dissonant non-motives at above-chance levels, choosing the motive at M=0.65 (SE=0.07), although this finding was not statistically significant from chance.13

Difference scores were also calculated for the forced-choice tests. Difference scores were determined by subtracting the listener's score before familiarization (taken from the pre-exposure forced-choice test) from the listener's score after familiarization (from the post-exposure forced-choice test). D-scores would be closer to 1 if listeners chose the motive after familiarization but not before familiarization, and d-scores would be at zero if they chose the motive equally before and after familiarization. All d-score means for the forced-choice test conditions were positive, which suggests significant learning of the trichordal motive despite any biases for these melodies prior to experimentation. The following conditions had d-scores significantly above chance (0.00): motives vs. small consonant non-motives (d-score=0.90, SE=0.05, $t(11)=18.57$, p<0.001, paired, two-tailed) and motives vs. large consonant non-motives (d-score=0.71, SE=0.09, $t(11)=7.93$, p<0.001, paired, two-tailed). Two other conditions were not significantly above chance: motives vs. large dissonant non-motives (d-score=0.19, SE=0.12) and motives vs. near-motives (d-score=0.02, SE=0.08).

3.3. Discussion

Findings suggest that uninformed musicians (without any formal training in twelve-tone music theory or information about what they were hearing) implicitly learned the common interval patterns within a performance of a novel twelve-tone composition. Listeners demonstrated this ability despite the challenges of hearing real music during familiarization, such as variable rhythms, changes of timbre, simultaneities, etc.

Reponses from the post-exposure forced-choice test suggest that listeners learned the most common interval pair in Op. 24/iii (the trichord containing intervals 8 and 11): listeners chose the motive over all other lures at above-chance levels. While participants selected motives over large dissonant non-motives at above-chance levels, it was not a significant finding. However, perhaps more familiarization was necessary in this condition: after hearing another familiarization phase, listeners went on to rate the motive significantly higher than large dissonant non-motives in the post-exposure probe-melody test.

Responses from the post-exposure probe-melody test also suggest that listeners learned the most common intervals heard during familiarization. Ratings for the trichordal motive were significantly higher than ratings for all other trichords (near-motives, small consonant non-motives, large consonant non-motives, and large dissonant non-motives). Melodic intervals heard often during familiarization, such as intervals +/- 11 and 13, were rated significantly higher than other similar, large, and dissonant intervals not heard during familiarization, such as intervals +/- 10 and 14.

Ratings of trichords from the post-exposure probe-melody test (and d-scores) strongly correlated with the distribution of interval pairs heard during a novel twelve-tone piece of music. This supports an extension of Krumhansl's research: whereas Krumhansl (1990, 75) has asserted that we attune to the distribution of pitch when hearing music with a hierarchical distribution of pitch, this study provides empirical evidence that we also have the capacity to learn the distribution of intervals (and interval pairs) when that pitch distribution is mostly flat and instead the music is predominated with certain intervals.

A possible confound with the experimental design is that listeners were tested on the stimuli before and after familiarization, and thus some learning may have occurred from the pre-exposure tests. However, these pre-tests provided necessary information about the bias listeners had for intervals prior to familiarization.

---

12 D-scores for motives were significantly higher than small consonant non-motives ($t(11)=9.18$, $p<0.0001$), large consonant non-motives ($t(11)=7.46$, $p<0.0001$), large dissonant non-motives ($t(11)=2.32$, $p<0.05$), and near-motives ($t(11)=2.90$, $p<0.05$), as determined by paired, two-tailed t-tests.

13 Not only were listeners able to discriminate the often-occurring motive during familiarization, but they also learned the near-motives. This demonstrates a fine-tuned ability to learn the distribution of interval pairs in a novel composition. Listeners discriminated between near-motives over large consonant non-motives at M=0.69 (SE=0.06) at above-chance levels ($t(11)=3.38$, $p<0.01$) and they discriminated near-motives over large dissonant non-motives at M=0.60 (SE=0.05) at above-chance levels ($t(11)=2.16$, $p=0.05$).

4. ANALYTICAL IMPLICATIONS

The experimental findings support an interval-based analytical approach to Op. 24/iii. This analysis provides ways to interpret the roles of the adjacent intervals within surface trichords, and to use this perceptual lens to ascertain the movement's formal structure.

Previous analyses of Op. 24/iii have described its twelve-tone aspects, such as that the first twelve notes are a row beginning on F, followed by a retrograde inversion of that row (e.g. Bailey, 1991). Other analyses have noted that every trichord in the movement is a member of SC[014], a byproduct of the fact that Webern uses a derived row in this composition (e.g. Babbitt, 1955). An analysis informed by cognitive research may instead focus upon the frozen intervals heard throughout the movement. Given the experimental findings above, the following observations of Op. 24/iii may be more perceptually salient than previous analytical approaches.

First, an analysis of the intervals in Op. 24/iii reveals that intervals 8 and 11 (heard in opposite directions) are pervasive on the musical surface. The listener can group the four intervallic versions of trichords heard often in the movement (intervals +8-11, -8+11, -11+8, and +11-8) into one understanding of the motive, which is defined in this analysis by its interval content. Indeed, Webern himself describes the motive as the smallest independent part of a musical idea, recognizable through repetition (Webern, 1963, 25-6).

We might also study the roles that the near-motives play in the composition. For example, some near-motives are only heard as chords (as opposed to melodies). Other near-motives are heard cadentially. For example, the near-motive with intervals 4 and 11 (such as -4-11 or +4+11) is used at the end of musical phrases in the beginning of the movement, such as in mm. 12-13. These cadential moments may stand out to listeners more because the contour of this near-motive is different from the contour of the motive (which contained one ascending and one descending interval).

Finally, the analytical listener may also attune to the motive to determine the formal divisions within Op. 24/iii. This type of listening would reveal that there are five sections in the composition, which agrees with the formal analyses of the movement by Bailey (1991) and Gauldin (1977). To illustrate: in the first section of the movement (mm. 1-13), melodic trichords predominate the musical surface, and most of these trichords are the motive. In Section 2 (mm. 14-28), trichords are still prominent on the musical surface. However, Section 2 differs from Section 1 because chords are now heard for the first time, and they are all near-motives. In Section 3 (mm. 28-41), surface trichords are obscured, and dyads instead pervade the musical surface. In the final two sections of the movement, trichords return to the musical surface and the motive becomes increasingly more common as the work comes to a close. In Section 4 (mm. 41-55), every note in the piano part is a member of a harmonic motive. In Section 5 (mm. 56-end), all notes are members of motives, and this motive is heard both melodically and harmonically.

Experimental findings described herein suggest that this interpretation of the movement is perceptually salient, and listening to intervals provides us with multiple ways of analyzing the composition: we can first attune to intervals to learn the motive. We can then use this motive to determine formal divisions in the composition, and we can even apprehend the roles of near-motives.

5. FUTURE DIRECTIONS

5.1. Analytical research

The interval-based analytical approach can be extended by studying the use of frozen intervals in other atonal and twelve-tone works. Indeed, some of Webern's other compositions also feature the "skewed octave" (pitch intervals 11 and/or 13), such as Op. 5/iii, Op. 25/i, Op. 27/ii, and Op. 27/iii). Listeners may hear meaning in these compositions by attuning to the prominent trichords containing these intervals, and such analyses may provide a narrative that aligns well with listeners' perceptions.

5.2. Experimental research

Future experimentation may test non-musicians to determine whether these findings are generalizable to larger populations. It also remains to be seen whether participants have the ability to learn the distribution of intervals when hearing other musical genres, especially when the distribution of pitch is not flat.

6. CONCLUSION

Experimental findings suggest that musician listeners can hear structure in twelve-tone music by attuning to intervals occurring on the musical surface. Participants implicitly learned the distribution of intervals given a recording of a novel piece of twelve-tone music containing recurrent intervals. These results support a listening strategy for a musical genre notoriously difficult to aurally comprehend, and this interval-based approach to twelve-tone music has both analytical and pedagogical applications.

7. REFERENCES

Alegant, B. (2001). Cross-partitions as harmony and voice leading in twelve-tone music. Music Theory Spectrum, 23*(1), 1-40.*

Babbitt, M. (1955). Some aspects of twelve-tone composition. The Score and IMA Magazine, 12, *53-61.*

Bailey, K. (1983). Webern's opus 21: Creativity in tradition. The Journal of Musicology, 2*(2), 184-195.*

———. *(1991).* The twelve-note music of Anton Webern: Old forms in a new language. *Cambridge and New York: Cambridge University Press.*

———. *(1996a). Symmetry as nemesis: Webern and the first movement of the Concerto, Op. 24.* Journal of Music Theory, 40*(2), 245-310.*

———. *(1996b).* Webern studies. *Cambridge and New York: Cambridge University Press.*

Brown, J. L. (2014). Cognitive models of interval perception in twelve-tone music: Hearing Webern's Concerto, op. 24, and other works. Ph.D. diss., University of Rochester.

Bruner, C. L. (1984). The perception of contemporary pitch structures. Music Perception, 2, *25-39.*

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. Journal of Experimental Psychology: General, 113, *394-412*.
Dienes, Z. & Longuet-Higgins C. (2004). Can musical transformations be implicitly learned? Cognitive Science: A Multidisciplinary Journal, 28, *531-558*.
Gauldin, R. (1977). The magic squares of the third movement of Webern's Concerto Op. 24. In Theory Only, 2*(11-12), 32-42*.
Headlam, D. (2012). The shape of things to come? Seeking the manifold attractions of tonality. [Review of the book A geometry of music: Harmony and counterpoint in the extended common practice, *by D. Tymoczko]*. Music Theory Spectrum, 34*(1), 123-143*.
Huron, D. (2006). Sweet anticipation. *Cambridge, MA: MIT Press.*
Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and the West. Music Perception, 2, *131-165*.
Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. Journal of Music Theory, 27, *75-97*.
Krumhansl, C. L. (1990). Cognitive foundations of musical pitch. *New York, NY: Oxford University Press.*
———. (1991). Memory for musical surface. Memory & Cognition, 19*(4), 401-411*.
Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a special representation of musical keys. Psychological Review, 89, *334-368.*
Krumhansl, C. L., Sandell, G. J., & Sergeant, D. C. (1987). The perception of tone hierarchies and mirror forms in twelve-tone serial music. Music Perception, 5, *31-78.*
Kuusi, T. (2003). The role of set-class identity in the estimation of chords. Music Theory Online, 9*(3).*
Lerdahl, F., & Jackendoff, R. (1983). A generative theory of tonal music. *Cambridge, MA: MIT Press.*
Meyer, L. (1967). Music, the arts, and ideas: Patterns and predictions in twentieth-century culture*. Chicago, IL: University of Chicago Press.*
Ockelford, A. and Sergeant, D. (2013). Musical expectancy in atonal contexts: Musicians' perception of "antistructure." Psychology of Music, 41*(2), 139-174*.
Samplaski, A. (2004). The relative perceptual salience of Tn and TnI. Music Perception, 21*(4), 545-559.*
Webern, A. (1963). The path to the new music, *ed. Willi Reich, trans. Leo Black. Bryn Mawr: Theodore Presser Co., in association with Universal Edition. Reprinted London: Universal Edition.*
———. (1991). Concerto for flute, oboe, clarinet, horn, trumpet, trombone, violin, viola, and piano, Op. 24, iii. Sehr langsam. [Recorded by Pierre Boulez[. On Webern: Complete works opp. 1-31 *[CD]. Holland: Sony Classical: S3K 45845. 1978 CBS Records.*
Westergaard, P. (1963). Webern and "total organization": An analysis of the second movement of Piano Variations, op. 27. Perspectives of New Music, 1*(2), 107-120*.
Youngblood, J. E. (1958). Style as information. Journal of Music Theory, 2, *24-35*.

SPACIOUSNESS OR EVENNESS? THE EXAMINATION OF IC-SPACE & HARMONIC DENSITY IN ATONAL MSUIC

Yi-Cheng Daniel Wu[1]
[1]*Soochow University School of Music, China*
Time: 17:30

INTRODUCTION

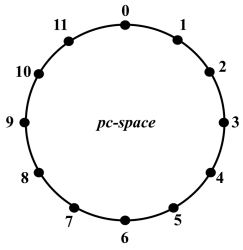
To make 20th- and 21st-century harmonic vocabulary more accessible, this article proposes a theory that considers the space of all ics and further derives a simple arithmetic formula examining the density of a chord in terms of its *degree of compactness*. In post-tonal music, there are four different methods that can categorize all intervals. They are ordered pitch interval, unordered pitch interval, ordered pitch-class interval, and unordered pitch-class interval. While the first two appear in pitch space, the last two appear in pitch-class space. This paper privileges the *unordered pitch-class interval*, which contains seven members of ics0−6. Two important strengths reinforce the perception of an interval in terms of its corresponding ic.

First, it allows us to classify pcset that are transpositionally or inversionally related into the same sc, for we generally believe that they sound alike based on their identical ic vector. Second, continuing with the first strength, many contemporary theorists use ic vector as a criterion to derive various methods of comparing sound similarity among different scs. These include Robert Morris's SIM, Eric Isaacson's IcVSIM, and Michael Buchler's SATSIM, among others.14 Thus, recognizing an interval in terms of its corresponding ic is crucial in music set theory applied by most theorists to analyze post-tonal music. For these reasons, my paper only considers the spectrum of ics, studying how they differ from one another by changing their *sizes*, *spaces* within this spectrum, while the other three types of interval perception will be left relatively unexplored.

Grounded in the spatial concept of all ics, I propose a simple, theoretical premise that interprets the sound of a post-tonal harmony. That is, if a chord is predominately composed of compact ics, it sounds more compact. Contrarily, if a chord is predominately composed of spacious ics, it sounds more spacious. To accurately compare the proportion between compact and spacious ics within a chord, I derive a measurement called the *average-VPticset*, which examines the density— the degree of compactness— of a chord. A resultant is always a number ranging from 0.00 to 6.00— from the most compact to the most spacious harmony. Importantly, we can use the

14 For a detailed summary and comparison of different sc similarity measurements, see Eric Isaacson 1990 and Michael Buchler 2001, among others.

average-VPticsets as references to access the relationship between any two post-tonal harmonies, for one may sound alike, or more compact/spatial than the other.

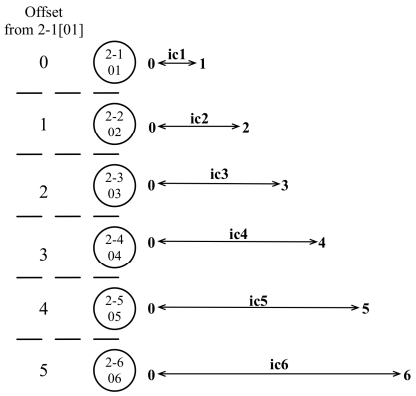


Ex. 1: Using a clock to represent pc-space

1. STRAUS'S OFFSET NUMBER

Ex. 1 shows a clock image, a schema used by most theorists to represent pc space. The points on the circle are the twelve pcs0–11. Mapping one pc to another via the shortest distance around the circle derives an ic, which will range from 0 to 6: from two pcs staying in the same position to those appearing on the opposite sides of the clock; or from two pcs that have no space in between to those that are farthest apart. This observation nicely supports our common understanding of the space of all ics. That is, ics0 and 6, respectively, project the most compact and the most spacious ics. However, based on recent research conducted by Joseph Straus, his theoretical viewpoint has inspired me to contemplate and re-conceptualize the space of all ics. Before introducing my theory, let me briefly discuss Straus's method. During this discussion, I point out the strength and a limitation in his research; I then use that limitation as a point of departure, proposing a new approach that can more relevantly interpret and compare the space of all ics.

In his 2005b article "Voice Leading in Set-Class Space," Straus extends his *offset numbers* derived from *fuzzy transformational voice leading* (1997, 2003, and 2005a) to describe the relative degree of *chromaticness* among different ics. His theoretical premise is that the offset number will increase and decrease its value based on the density of the pitch classes (pcs) within a sc. The smaller the offset number, the higher the density. He uses six different terms to define the density of a chord. They are "compactness, denseness, chromaticness" for scs with maximally packed pcs, and "dispersion, spaciousness, evenness" for scs with maximally spread pcs (2005b, 67). The left side of Ex. 2 (offset numbers and scs) shows Straus's analysis of the six dyadic scs from sc2-1 to sc2-6. He defines sc2-6 as representing the maximally spacious and even dyadic sc for it receives the largest offset number. If we observe the space between the two pcs in sc2-6 (see the double arrows on the right side of Ex. 2), it is ic6, which is indeed more spacious than the other five ics1–5 formed respectively by sc2-1 to sc2-5. However, once I consider the evenness and compare the six ics1–6, I cannot help but wonder: why and how is ic6 more even than ics1–5? Clearly, ic6 stretches out into the most expanded space, one that is spacious enough to cover ics1–5. But based on the image presented in this example, we can hardly picture ic6 as being both maximally spacious and even simultaneously. Now, we are encountering the issue that questions the relationship between Straus's evenness and his spaciousness—how can we define ic6 as projecting a maximally even ic?



Ex. 2: Compare the spaces of ics1–6

2. RE-CONCEPTUALIZE THE SPACE OF AN IC

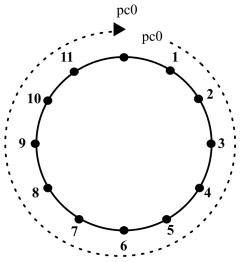
To perceive ic6 as the maximally even ic, we must reference it to the context of an octave, for ic6 (the tritone) can divide an octave into two *even* halves. Thus, only when the octave and ic6 exist simultaneously can we perceive the quality of evenness. In other words, the sense of evenness must rely on the reference of an octave, whose ic representative is ic0. Notice that traditionally ic0 represents both

octave and its inversion of the unison. Here, to satisfy the condition that ic0 always projects the maximally spacious ic, ic0 will only ever refers to an octave. With this particular condition considered, I must declare that in the theoretical spectrum of my paper the octave is not inversionally equivalent to a unison, and we must hear unison as a "tone repetition," which does not form an interval.

Within this spatial environment of ic0, ic6 represents evenness, and ic1 represents compactness. Based on this structure, I fill in the other four ics2–5 according to their sizes from small to large, and the result is shown in Ex. 2. There are two advantages in this new ic-ordering from ic1 to ic0 based on my refined version of Straus's harmonic definitions. First, none of the ics has multiple or overlapping descriptions of harmonic quality. Second, while the qualities of chromaticness and spaciousness— ones that are strongly associated with the "size" of a space— consistently appear on the polar sides in Ex. 3, the quality of evenness, which is less associated with the "size" of a space, embeds within that environment. Thus, we can perceive a sense of the gradual expansion or contraction in space created by ics moving away or towards ic1, while the sense of the evenness is understood in a more referential fashion with regard to ic0.15

	compact					even	spacious
ics	1	2	3	4	5	6	0 octave

Ex. 3: Ordering the seven ics based on their spaces



Ex. 4: Using a clock to represent pc-space

Moreover, in order to strengthen the concept of ic0 as always being the most spacious ic, I slightly modify the traditional measurement of an ic in the clock schema in Ex. 1. I set up two rules that must be followed to measure an ic. First and foremost, no pc ever remains stationary. Second, we only consider the shortest distance between two pcs. Thus, if we map a pc onto itself, the shortest distance will be a full-circle progression, which corresponds to an ic0 and creates the largest space. For instance in Ex. 4, to map one pc0 onto another by following the above two rules, the first pc0 cannot stay; it must move in order to get to the second pc0. If we need to move the first pc0, what is the most efficient way to move it to the next pc0? Apparently, the shortest distance and the most efficient way is to have it circle around the entire clock to return to its original point— the second pc0 (see the dashed arrow directing pc0 to move around the clock in Ex. 4). This full-circle progression, then, projects the largest space of ic0 within this clock. Next, if we move one pc to another, we will derive the remaining six ics1–6, those that have relatively smaller spaces than that of ic0. As a result, ic0 and ic1 now turn out to be the most spacious and the most compact ics respectively, which are exactly the same as those laid out in Ex. 3.

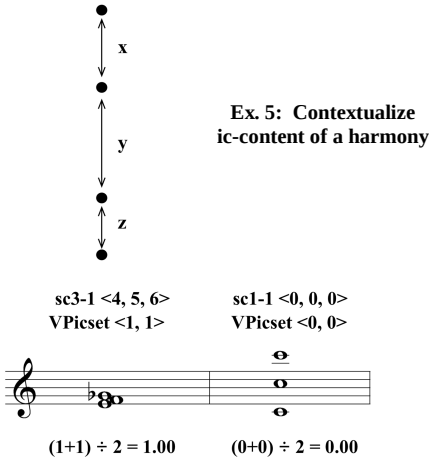
3. HARMONIC DENSITY— THE DEGREE OF COMPACTNESS

Based on this new spatial interpretation of all seven ics, I propose a measurement to study the harmonic density of a chord— i.e., the degree of compactness.16 Ex. 5 demonstrates my measurement. The ic is formed by each pair of the adjacent voices, which is <z, y, x> (ordered from low to high). I call this ordered ic-set the voice-pair ic set— or, the VPicset. My theoretical premise is that if these three ics <z, y, x> predominately appear to the extreme left in Ex. 3, they make a compact chord with high harmonic density. Contrarily, if they predominately appear to the extreme right in Ex. 3, they make a spacious chord with low harmonic density. To more accurately define how compact a harmony is, I propose a method that sums up all these ics and then averages the total. The derived result is a mean number called the *average-VPicset*, which reflects the particular degree of density of a chord.17 The smaller the average-VPicset, the higher density the harmony. There is, however, a crucial problem in this harmonic measurement. It works if and only if there is no ic0.

For instance, there are two chords in Ex. 6— sc3-1 and sc1-1. Comparing their VPicsets, all two ics in sc1-1, <0, 0>, are more spacious than those in sc3-1, <1, 1>, since both ic0 appears to the right of ic1 in Ex. 3. Summing up the corresponding ics and averaging the total derives the average-VPicsets 1.00 (sc3-1) and 0.00 (sc1-1). Based on the above notion of "the smaller the average-VPicset, the higher density the harmony," my result suggests that sc1-1 with the number 0.00 is more compact than sc3-1, whose average-VPicset is 1.00. However, the image of sc1-1 being more compact than sc3-1 actually conflicts with my earlier observation, which reveals that all the ics in sc1-1 are more spacious than those in sc3-1. This conflict between my observation and the resultant average-VPicsets derives

15 The detailed and classic study of the evenness appears in John Clough and Jack Douthett's 1991.
16 Notice that I use the term "compactness" instead of "chromaticness" to differentiate my theory from Straus's.
17 In Ex. 5, the average-VPicset is (z+y+x)÷3.

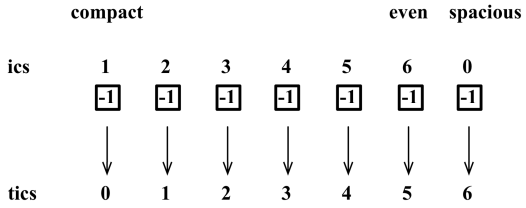
from the *exact value of the number*. Although the number 0 represents the most spacious ic in Ex. 3, its exact value is actually smaller than the remaining ics1 to 6. Thus, in the process of averaging the sum of the ics, zero will decrease, not increase, the value of an average-VPicset. Apparently, using the number 0 to represent the most spacious ic turns out to be irrelevant for this harmonic measurement, and, most importantly, the final result cannot satisfy the notion of "the smaller the average-VPicset, the higher density the harmony."



Ex. 5: Contextualize ic-content of a harmony

Ex. 6: Average-VPicsets of sc3-1 and sc1-1

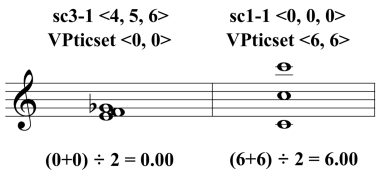
To solve the problem illustrated in Ex. 6, we must conceive of a way that can better represent the spatial value of the seven ics in Ex. 3. The solution to this problem is to re-value all the ics by subtracting one degree from each ic. As a result, ic1 equals 0, ic2 equals 1, …, ic6 equals 5. But what about ic0? Since the seven sequential ics are repetitive and form a closed group, these new revalued numbers must be sequential and repetitive as well. If subtracting one degree from the first ic1 derives the first number of 0, the next six sequential numbers must range from 1 to 6. As a result, the last ic0 corresponds to the number of 6, which represents the most spacious ic among all the seven ics (see Ex. 7).



Ex. 7: Transformed ics

Observing these new numerical representations, the most compact ic1 has the smallest value of 0, while the largest number 6 describes the most spacious ic0. I call these revalued ics the "transformed ics" (tics). The greatest strength of these new revalued numbers is found in their ability to provide a more representative picture of "the smaller the tics, the more compact the ics." Correspondingly, several tics form a set called VPticset. Similar to the average-VPicset, the average of a VPticset will be a number (average- VPticset) that accommodates the density of a chord, which ranges from 0.00 to 6.00.18 Within this range, the smaller the average-VPticset, the higher density the harmony. If we use tics to replace ics in Ex. 6, the average-VPticsets are 0.00 (sc3-1) and 6.00 (sc1-1, see Ex. 8). Compared to the earlier result in Ex. 6, these two new numbers better interpret the harmonic density of sc3-1 and sc1-1: three adjacent chromatic pcs are more compact than two octaves. Thus, in order to accurately compare the harmonic degrees of compactness, all ics must be translated into tics before any further calculations.

18 If a VPticset is only composed of the most compact tic0, summing up all the tics and averaging the total derives the average-VPticset of 0.00, which corresponds to the smallest number and represents the most compact harmony. Contrarily, if a VPticset is only composed of the most spacious tic6, summing up all the tics and averaging the total derives the average-VPticset of 6.00, which corresponds to the largest number and represents the most spacious harmony.

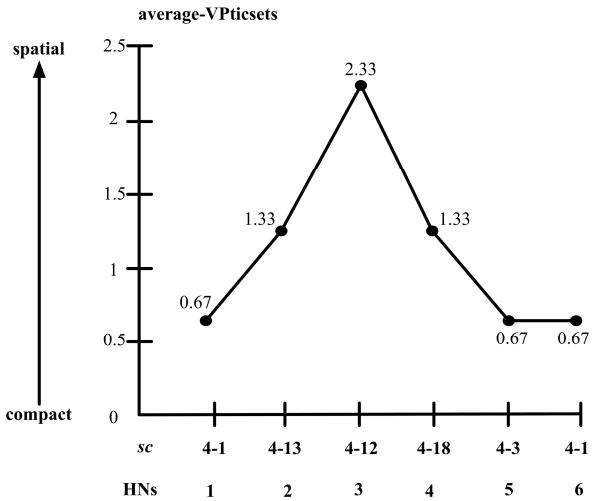


Ex. 8: Average-VPticsets of sc3-1 and sc1-1

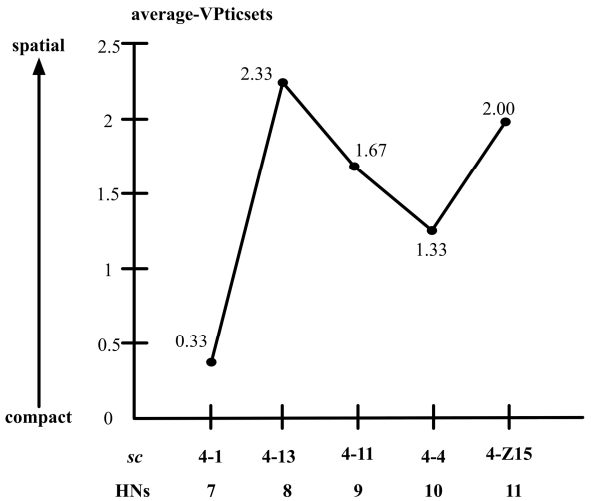
4. ANALYSIS— CRAWFORD'S STRING QUARTET, MVT. III

To understand the practical advantage of this harmonic measurement, I conclude my paper with an analysis of two passages from the third movement of Ruth Crawford's String Quartet (see Exs. 9 and 10). On the surface level, these two passages share several common contextual features, which suggest that they project a similar musical progression. They both begin with sc4-1 followed by sc4-13. These two scs are all articulated by the meter of $\frac{4}{4}$.19 Although their final tetrachords are not the same— one is sc4-1 (Ex. 9) and the other is sc4-Z15 (Ex. 10), they both, nevertheless, last for the long duration of four measures, creating a static, frozen, and stable motion that suggests a sense of a closure. But, besides these common contextual features, one must ask if there is a more fundamental element supporting the similarity between these two passages. Since my topic focuses on the density of a harmony, I apply the average-VPticset to analyze the chords in these two passages, and then use my findings as a means of revealing a deeper similarity, or perhaps difference, between Exs. 9 and 10.

To easily identify these harmonies from one another in these two examples, I assign each chord with a particular harmony number (HN, see the top row of both examples). Meanwhile, to represent these average-VPticsets in a way that mimics the flow of harmonies, I use Graphs 1 and 2 to respectively illustrate the relative degrees of the average-VPticsets in Exs. 9 and 10. The vertical axis in the graphs marks the average-VPticsets from 0.00 to 2.50. The sc names and HNs, in turn, appear on the horizontal axis. The dots, mid-graph, are the average-VPticsets of their corresponding tetrachords. Connecting the successive dots in one graph forms a figure that represents the harmonic progression for its corresponding example.



Graph 1: Crawford, String Quartet, Mvt. III, mm. 20–29; graphical representation of the average-VPticsets



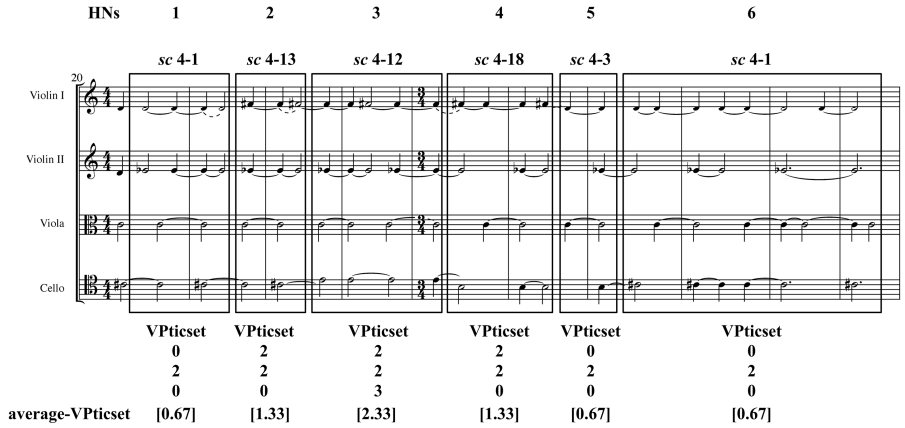
Graph 2: Crawford, String Quartet, Mvt. III, mm. 49–56; graphical representation of the average-VPticsets

Before any further discussion, I must define that within each graph the lowest dot (or the smallest average-VPticset) represents the most compact chord, and the highest dot (or the largest average-VPticset) represents the most spacious chord. The dots between these two extremes represent mediate chords. Comparing these two graphs, Graph 1 projects an upward symmetrical arch. The first half of the arch forms an ascending line, which begins with the most compact chord (HN 1) and gradually rises to the most spacious chord (HN 3) by passing through a mediate chord (HN 2). However, reaching the higher range of the graph is only temporary, because just after the ascent to HN 3, the arch immediately inverts this ascending line and creates a symmetrical descent, which begins with the most spacious chord (HN 3) and gradually descends to the most compact chord (HN 5) by passing through a mediate chord (HN 4). After HN 5, the arch maintains its motion by progressing to one more chord (HN 6), whose average-VPticset is the same as that of HN 5. Moving to Graph 2, it projects a markedly different figure from that of Graph 1, which forms a wave. It begins with a significantly direct ascent from a dot in the lowest range of the graph (HN 7, the most compact) to that in the highest range (HN 8, the most spacious). As soon as

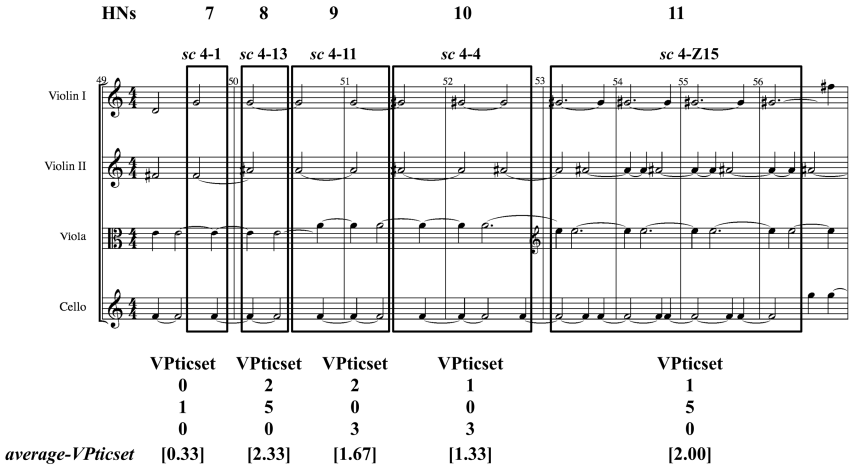
19 Crawford repeatedly alternates the meters of $\frac{3}{4}$, $\frac{4}{4}$, and $\frac{5}{4}$ in this movement.

the wave arrives at HN 8, the remainder of its progression becomes comparatively smoother, as it moderately descends to HNs 9 and 10 and then slightly rises again to HN 11.

Therefore, through the above detailed analyses of Graphs 1–2, we understand that each graph contains a particular figure, which projects a markedly unique harmonic progression for its associated musical passage. In the first passage (Graph 1 and Ex. 9), Crawford distributes the most compact (HNs 1, 5, and 6) and the most spacious chords (HN 3) over the entire harmonic progression, forming a gradual ascent and descent that covers the other two mediate chords (HNs 2 and 4). In contrast, in the second passage (Graph 2 and Ex. 10) Crawford uses a different harmonic progression, where the most compact and spacious chords are adjacent to each other (HNs 7 and 8), forming a considerable ascent at the very beginning of the passage. In addition, unlike the arch in Graph 1, this ascent is not deviated or interrupted by any other chords with rather mediate degrees of compactness, because they (HNs 9–11) all appear only after the initial ascent in Graph 2. According to the observation and comparison between Graphs 1 and 2, I conclude that although Crawford uses several common contextual features to compose these two passages, she nonetheless carefully and deliberately varies their bases by superimposing them on two different harmonic progressions.20



Ex. 9: Crawford, String Quartet, Mvt. III, mm. 20–29; VPticsets and average-VPticsets



Ex. 10: Crawford, String Quartet, Mvt. III, mm. 49–56; VPticsets and average-VPticsets

20 It is worth mention that voice crossings constantly appear throughout this movement in Crawford's String Quartet. Ellie Hisama uses the term "degree of twist" to define the number of voice crossing within a chord in this particular movement (1995, 298). For instance, the first sc4-1 in Ex. 9 has a degree of twist 2— Vc over Vla and Vln II over Vln I. Based on her analysis of the degree of twist, Hisama discovers multiple climaxes achieved by the gradual progressions moving toward and then away from an exceedingly twisted texture, which do not coincide with the registral and dynamic climax at m. 75. For Hisama, the highly twisted texture represents what she believes to be a "feminist" climax (1995, 305), which is opposed to the register and dynamics representing a more contextual, traditional, and "masculine" form of climax. (Also, Edward Gollin [2009] extends Hisama's study by applying a *Cayley graph* to analyze how the registral permutations of the four string instruments create a transformational network, and further uses his results to support Hisama's view of feminist climax in Crawford's String Quartet, Mvt. III.)

5. CONCLUSION

The above analyses illustrate the practical advantage of the average-VPticset as realized in the compositions by Crawford. Finally, I would like to conclude my paper with a brief discussion of other potential strengths in this harmonic measurement. To a certain extent, comparing two chords based on their resultant average-VPticsets is arguably reminiscent of other contemporary harmonic similarity measurements, for they all consider the ic instead of other types of intervals in developing various theories. But what are the differences between this and other theories? I summarize three strengths explaining why I believe my measurement represents a more applicable and useful method. First, its arithmetic formula is more mathematically friendly, composed of only two easy steps— sum up all tics and then average the total. Second, it accommodates scs with either identical or different cardinalities. Third, instead of using the general words "similar" and "dissimilar" to describe the relationship between any two chords, I assign each one the more concrete meaning of "compact" and "spacious." With this additional meaning, we obtain a more solid and clearer sense of the (dis)similarity of any two chords— because they are alike or different in terms of their harmonic densities. Most importantly, it opens up a new and alternative approach for us to understand post-tonal harmonies, thus potentially making them more accessible to those who seek to understand the diverse contextual styles of so many 20th- and 21st-composers.

6. REFERENCES

Buchler, Michael (2001). Relative Saturation of Interval and Set Classes: A New Model for Understanding Pcset Complementation and Resemblance. Journal of Music Theory *45: 263–343.*

Clough, John, and Douthett, Jack (1991). Maximally Even Sets. Journal of Music Theory *35: 93–173.*

Gollin, Edward (2009). Form, Transformation and Climax in Ruth Crawford Seeger's String Quartet, Mvmt. 3. Mathematics and Computation in Music *37: 340–46.*

Hisama, Ellie (1995). The Question of Climax in Ruth Crawford's String Quartet, Mvt. 3. In Concert Music, Rock, and Jazz since 1945, *eds. Elizabeth West Marvin and Richard Hermann, 285–312. Rochester: University of Rochester Press.*

Isaacson, Eric (1990). "Similarity of Interval-Class Content between Pitch-Class Sets: The IcVSIM Relation. Journal of Music Theory *34: 1–29.*

Straus, Joseph (1997). Voice Leading in Atonal Music. In Music Theory in Concept and Practice, *eds. James M. Baker, David W. Beach, and Jonathan W. Bernard, 237–74. Rochester: University of Rochester Press.*

___ *(2003). Uniformity, Balance and Smoothness in Atonal Voice Leading.* Music Theory Spectrum *25: 305–52.*

___ *(2005a). Voice Leading and Atonal Pitch Space. In* Introduction to Post-Tonal Music, *3rd ed, 107–12. Upper Saddle River, NJ.: Pearson Prentice Hall.*

___ *(2005b). Voice Leading in Set-Class Space.* Journal of Music Theory *49: 45–108.*

Wu, Yi-Cheng Daniel (2012). Reflection and Representation: A Unitary Theory of Voice Leading and Musical Contour in 20th-Century Atonal and Serial Contrapuntal Music. PhD diss., University at Buffalo.

[5B] Musical Schemata and Cognition

KYOTO 16:30- 18:00 Tuesday 5 2014

A COGNITIVE APPROACH TO MUSICAL GESTURE

Hye-Yoon Chung[1]
[1]*Myongji University, Korea, Republic of*

Time: 16:30

Today, under the strong force of the logical positivism which has dominated the music scholarship since the late 20th century, the description of music as gesture is inclined to be disregarded as a rhetorical expression of something else and therefore as replaceable with more literal and strict statement. Behind this tendency, there is an unproved belief that gestural description of music is unfounded and unreliable. This paper argues that gestural description of music is a linguistic reflection of gestural understanding of music which is an essential part of our musical experience. In order to vindicate the gestural understanding of music, this paper elucidates the ultimate foundation of the experience of music as gesture from a cognitive approach through the concept of the image schemata which was first suggested by Mark Johnson and George Lakoff(1980) in the attempt to reveal the bodily basis of our ordinary understanding. For this purpose, it first presents the definitional haracteristics of the image schemata, and then shows the essential features of musical gesture. Then, it analyzes the gestural experience of music through the concept of the image schemata. Through this, it is to be shown that the characteristics of the image schemata correspond to those of musical gesture in a significant way. First, the mental substantiality of musical gesture as an intentional object in the endo-somatic world resonates with the nature of the image schemata as a mental pattern shaped through a repeated bodily movement. Second, musical gesture and the image schemata have the same cognitive function as a path through which we can understand the given objects as meaningful to us. Third, the structural contour which the image schemata suggests as a highly refined Gestalt corresponds to the structural feature of musical gesture as an integrated Gestalt. Ultimately, this paper argues that listeners can experience music as gesture, since s/he perceives the sequence of sounds as a meaningful whole through the image schemata and that the gestural understanding of music at least as valid as the understanding of music in terms of the formal ideas and the technical vocabularies.

THE INSTITUTIONALIZATION OF APPRENTICESHIP IN THE GREAT CONSERVATORIES: A COGNITIVE INTERPRETATION OF A NON-VERBAL PRAXIS

Robert Gjerdingen[1]
[1]*Northwestern University, USA*
Time: 17:00

1. Background: Music, like the other crafts of pre-industrial Europe, was learned through apprenticeship. When formalized, an apprenticeship often involved a seven-year indenture of a young boy or girl to a master or mistress. Unpaid labor was exchanged for entry into the "mystery" of the craft. The four _conservatori_ or foundling homes of Naples were the first to institutionalize this indenture, ultimately aggregating as many as 600 boys for the intense study of music. Just as London's four Inns of Court raised clerkship to a higher level in English law, so the conservatories of Naples created a critical mass of young talent such as Europe had never seen. By the end of the nineteenth century almost every industrializing nation had established its own version of a music conservatory. In the last decade there has been a revolution in how scholars interpret the training at these conservatories (Gjerdingen 2007; Sanguinetti 2012). Exercises previously decried as mindless and devoid of any tonal theory (Dehn 1840) can now be understood as sophisticated guides to the development of what contemporary linguists term a usage-based or construction grammar (Bybee 2006; Goldberg 2006). The students became fluent "speakers" of a polyphonic language. The two most famous of these students today would be Debussy and Ravel.

2. Aims: The paper will detail how harmony was taught at the Paris Conservatory during the nineteenth century, stressing the cognitive development of the young students (typically ages 11–18). Several of the polyphonic constructions taught there will be illustrated from textbooks of the period.

3. Main contribution: The thousands of documents preserved by the Paris Conservatory can be read as a longitudinal study of the cognitive development of talented young musicians. In particular, the annual contests in harmony and counterpoint provide evidence of a type of Pavlovian conditioning. Students were given the stimulus of a single bass or melody and asked for a polyphonic response crafted from previously learned constructions.

4. Implications: Training at the Paris Conservatory challenges our notions of what "harmony" means, and suggests that statistical learning can play a major role in music instruction.

THE STRUCTURE OF COGNITIVE AND MUSICAL ABILITIES

Daniel Müllensiefen[1], *Amit Avron*[1], *Naoko Skiada*[1]
[1]*Department of Psychology, Goldsmiths, University of London, United Kingdom*
Time: 17:30

1. Background: Gardner (1983, 2006) formulated a theory of Multiple Intelligence that explicitly posits musical intelligence as a separate dimension along 8 other and independent types of intelligence. However, Gardner's model has found mixed empirical support in the literature but previous studies have commonly assessed only a narrow range of musical abilities. In addition, different mechanisms have been discussed (e.g. Schellenberg, 2011) to explain the relationship between musical training and musical abilities, intelligence and personal background variables.

2. Aims: Firstly, this study aims to replicate previous findings regarding the role of musical abilities with regards to the overall structure of intelligence using a novel musical assessment battery. Secondly, this study tests the fit of different causal mechanisms that have been suggested in the literature to explain the relationships between perceptual musical ability, musical training, general intelligence, and socio-economic status (SES).

3. Method: 130 adults were assessed for verbal and perceptual intelligence via the Wechsler Abbreviated Scale of Intelligence (2nd edition), as well as for their perceptual musical abilities and self-reported musical engagement via the Goldsmiths Musical Sophistication Index. Socio-economic status was measured using the NS-SEC inventory. The structural equation models (SEMs) were employed to compare the multiple intelligence model to hierarchical models of intelligence as well as to assess the potential influence of musical training and SES on musical abilities and g.

4. Results: Results very clearly suggest the presence of a general intelligence factor g that is significantly related to verbal and perceptual intelligence as well as perceptual musical abilities. Furthermore, the best model incorporating SES and musical training indicates a) a positive influence of SES on musical training, b) a positive influence of musical training on musical perceptual abilities, c) a positive influence of musical training on g and d) the absence of any effect of SES on g.

5. Conclusions: Results clearly suggest a positive relation between musical abilities and g in this adult sample. Musical training has a positive influence on both musical abilities and g. In turn socio-economic status appears as a driving influence on musical training but shows no direct effect on g.

[5C] Rhythm in Language, Music and Movement

MONTREAL 16:30- 18:00 Tuesday 5 2014

THE ACQUISITION OF SPEECH RHYTHM IN CHINESE-,ENGLISH-, AND JAPANESE-LEARNING INFANTS

Zhimin Bao[1], *Yuko Yamashita*[1], *Kazuo Ueda*[1], *Yoshitaka Nakajima*[1]
[1]*Kyushu University, Japan*

Time: 16:30

1. Background: Learning rhythm plays a crucial role in both music and language learning, as a rhythmic pattern provides the overall shape of a music piece or a spoken material for listeners. From the first year of life, infants are sensitive to rhythm.

2. Aims: The present study focused on the acquisition of the speech rhythm in infants and toddlers from different language backgrounds. Chinese, English , and Japanese are categorized into different linguistic-rhythm groups; a syllable-timed, a stress-timed, and a mora-timed language, respectively. The purpose of the present study was to explore the developmental changes of rhythm and spectral fluctuation in Chinese-, English- and Japanese-speaking infants.

3. Methods: Three age groups (15, 20, and 24 months) were selected, because infants diversify phonetic inventories with age. Natural speech of the infants was recorded in a quiet room in each infant's home for about 2 hours. For some infants, their speech was recorded more than once at different ages. Cepstral analysis was performed and the correlations between the power fluctuations of the critical-band outputs were represented by factor analysis indicating how the critical bands should be connected to each other, if a listener is to differentiate sounds in the infant's speech efficiently. Autocorrelation analyses followed in order to visualize speech rhythm.

4. Results: The present analysis identified three factors, as had also been observed in adult speech, at 15, 20, and 24 months of age in all three linguistic environments. These three factors were shifted to a higher frequency range, compared to those obtained from adult speech, corresponding to the smaller vocal-tract sizes of the infants. The results suggest that the vocal tract structures of the infants acquire an adult-like configuration between 15 and 24 months of age regardless of language environment. The autocorrelation analyses revealed difference between Japanese and the other languages in terms of rhythm.

5. Conclusions: The amount of utterances with periodic nature related to shorter time increased with age in the three different language environments. This trend was clearer in the Japanese environment. It seems important for systematizing linguistic education to examine how infants acquire speech rhythm at an earlier age.

NON-LINEAR CHANGES IN RHYTHMIC VARIABILITY OF EUROPEAN ART MUSIC: QUANTITATIVE SUPPORT FOR HISTORICAL MUSICOLOGY

Niels Chr. Hansen[1,2,3], Makiko Sadakata[4,5], Marcus Pearce[6]
[1]Music in the Brain, Center of Functionally Integrative Neuroscience, Aarhus University Hospital, Denmark, [2]Royal Academy of Music Aarhus, Denmark, [3]Department of Aesthetics and Communication, Aarhus University, Denmark, [4]Department of Musicology, University of Amsterdam, Netherlands, [5]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, [6]Centre for Digital Music and Centre for Research in Psychology, Queen Mary, University of London, United Kingdom

Time: 17:00

It is a long-held belief in historical musicology that the prosody of composers' native languages is reflected in the rhythmic and melodic properties of their music. Applying the normalised Pairwise Variability Index (nPVI) to speech alongside musical scores, research has established quantitative similarities between durational variability in language and music. This work capitalises on the fact that syllable-timed languages like Italian and French have low nPVI while stress-timed languages like German and Austro-German have high nPVI. Extending this approach to analyses of historical developments, a recent paper ascribed linearly increasing nPVI in Austro-German music, but not Italian music, to waning Italian influence on Austro-German music after the Baroque Era. This "Italian Influence Hypothesis" is, however, a post-hoc hypothesis, and since we cannot perform controlled experiments on historical data, replication with more sensitive methods and new repertoire is required. Turning to French music, we hypothesised an initial increase and a subsequent decrease, based on documented increasing German influence on French music after the Baroque and reported decreasing nPVI in French vocal music composed 1840-1900. In contrast to previous studies that only applied linear modelling, our prediction necessitated polynomial modelling to detect non-linear historical developments. Mean nPVIs were computed for 34 French composers (midpoint years: 1700-1941). Moreover, previous data were available for 21 Austro-German (1672-1929) and 15 Italian composers (1613-1928). Predicting mean nPVI from midpoint years, a 2nd-order polynomial outperformed a linear function for French composers. Adding another parameter did not improve this fit significantly. Linear analyses replicated decreasing nPVI specifically for composers born after 1820, and, furthermore, found a preceding increase with identical effect size to that previously reported for Austro-German composers. Previous findings for Austro-German (linear increase) and Italian composers (no change) were similarly replicated. Since French and Italian have similar linguistic nPVI, increasing French musical nPVI cannot represent decreasing influence from Italian music. Rather, this development can be explained in terms of a novel "Austro-German Influence Hypothesis". This is consistent with musicological observations of Austrian/German influences on classical French music up until the mid-19th century, after which French music diverged into an Austro-German school and a French nationalist school. In sum, using musical nPVI analysis, we provide quantitative support for music-historical descriptions of an Italian-dominated Baroque (composer birth years: 1600-1750), a Classical Era (1750-1820) with Austro-German centres of gravity (e.g. Mannheim, Vienna), and a Romantic Era (1820-1900) with greater national independence.

THE EFFECTS OF RHYTHMIC AUDITORY STIMULATION ON GAIT PARAMETERS OF ADOLESCENTS WITH TRAUMATIC BRAIN INJURY DEPENDING ON THE PRESENCE OF MELODY

Soo Ji Kim[1], Hye Ji Park[1]
[1]Graudate School of Education, Ewha Womans University, Korea, Republic of

Time: 17:30

The purpose of this study was to examine the effect of rhythmic auditory stimulation (RAS) on gait parameters of adolescents with traumatic brain injury (TBI) depending on the presence of melody. Three adolescents with TBI participated in this study and each participant received a total of twelve individual RAS training. Basic chord patterns without the melody or complex chord patterns with the melody were used during the training. At pre and posttest, temporal parameters including cadence and velocity and kinematic parameters via VICON 370 Motion Analysis System were measured. The results of this study demonstrated that there were no significant differences in the velocity depending on the type of the stimulation (i.e., chord patterns with or without melody) for all participants. All participants showed improvement in the gait function after RAS training. The cadence, velocity, stride length were increased and the stride time was reduced at posttest than at pretest. The index for symmetry was also enhanced. The motion analysis demonstrated that movement patterns of front pelvic tilt, hip joint and knee joint were closer to the criteria for the normal gait, which indicates that the walking stance became more stable. The research findings indicate that rhythm is the primary factor in mediating gait functions when applying RAS training. This study also showed that RAS training can effectively improve the gait function on adolescents with TBI, which has been investigated less extensively with regard to gait training. As an initial attempt to examine the effects of specific musical elements on gait measures in RAS training, this study suggests that further studies would be needed to confirm how to incorporate different levels of musical elements into RAS training to enhance its effects on gait function of this population.

[5D] Music and Memory 1

LIÈGE 16:30- 18:00 Tuesday 5 2014

CAPTURING EARWORMS IN THE LAB: EFFECTS OF COGNITIVE LOAD ON THE APPEARANCE OF INVOLUNTARY MUSICAL IMAGERY

Georgina Floridou[1], Victoria Williamson[2,3], Lauren Stewart[1]
[1]Department of Psychology, Goldsmiths, University of London, United Kingdom, [2]Hochschule Luzern – Musik, Lucerne University of Applied Sciences and Arts, Switzerland, [3]Department of Music, University of Sheffield, United Kingdom

Time: 16:30

1. Background: The term «Involuntary musical imagery», or 'earworms', refers to the ubiquitous experience of a musical snippet coming into the mind without effort and then repeating. Previous studies provide conflicting reports about the circumstances of INMI onset: some suggest that INMI is more likely to occur during low concurrent cognitive load while others show an increase in INMI during high cognitive load, relative to baseline. To date, the effect of systematically increasing cognitive load on INMI remains unknown.

2. Aims: The first aim of this study was to trial a new non-leading, implicit method for INMI induction and assessment. The second aim was to use this method to investigate the effect of concurrent cognitive load (four levels) on INMI occurrence. The final aim was to measure subsequent INMI experiences over a 24-hour period.

3. Method: One hundred and sixty people participated, making this the largest single blind INMI experiment to date. Participants watched two non-speech film trailers (one with instrumental music and one lyrical) and engaged in a film evaluation exercise. Subsequently, they completed one of four 5-minute tasks, which systematically increased in cognitive load. Finally, they completed a new 'mind activity questionnaire' that allows for implicit sampling of INMI experiences rather than direct leading assessment. The same questionnaire was completed again after 24 hours.

4. Results: INMI induction at silent baseline was 52.5%, a rate which is notably lower than that reported in previous studies that used explicit, leading assessments. This induction rate was reduced to 27.5% in low cognitive load conditions, and reduced again in both medium and high cognitive load conditions, which did not differ from each other (10% and 12.5%, respectively). Lyrical music occurred more as INMI compared to instrumental music. In the 24 hour follow-up, 21.4% of people experienced INMI.

5. Conclusions: This new implicit, non-leading response paradigm provides valuable methodological insights for future INMI work and has clarified the specific effects of differential, increasing cognitive load on INMI onset.

DIFFERENT MECHANISMS FOR RHYTHM MEMORY IN REPRODUCTION AND COMPARISON TASKS

Niall Klyn[1], Yong Jeon Cheong[1], Erin Allen[1], Udo Will[1]
[1]The Ohio State University, USA

Time: 17:00

ABSTRACT

A recent study Hung (2011) found significant differences in processing of vocal and instrumental rhythms in early auditory pathways in temporal lobe, and in a behavioral study we found differential processing of vocal and instrumental rhythms in short-term memorization (Klyn et al., submitted). Saito and Ishio (1998) suggested the involvement of the phonological loop in rhythm reproduction tasks. In the current study we examine whether the phonological loop is involved in the differential memorization of vocal and instrumental rhythms. Experiment 1 tests a possible involvement of the phonological loop in rehearsal strategies for rhythm memorization with a dual task paradigm. Subjects listened to stimulus pairs containing both vocal and clapstick rhythm patterns, and made same-different judgments on either the vocal or the instrumental rhythm of each pair while performing a concurrent motor task (sub-vocal articulation or finger-tapping). Results show that concurrent tasks affect both vocal and instrumental rhythms, but provide no evidence for phonological loop involvement in the delayed decision task. As these results differ from Saito and Ishio's (1998) findings, experiment 2 tests the

involvement of the phonological loop in a delayed rhythm reproduction task. Results show that a concurrent sub-vocal articulation task decreases memory performance significantly more than a concurrent tapping task and suggest the involvement of the articulatory loop in this task. Our experiments indicate that the phonological loop is not involved in all rhythm memory processes: it seems to be recruited only when we memorize rhythm for performance, not when rhythm is memorized for comparison.

1. INTRODUCTION

Research on auditory processing (e.g., Belin et al., 2000; Bent et al., 2006; Levy et al., 2001, 2003; Vouloumanos, 2001; Zatorre et al., 2002) has revealed differential processing of vocal and non-vocal sounds. Musical rhythm processing, on the other hand, is generally regarded as processing of abstract temporal features, and the question whether sounds from different sources lead to differential effects in rhythm processing has not been systematically investigated. However, Hung (2011) advanced evidence that even within the auditory modality rhythm processing is not independent of features determined by the sound source. Her study showed significant behavioral (reaction time, accuracy) and imaging (fMRI) differences for processing of vocal and instrumental rhythms. In a recent same-different decision study on vocal and instrumental rhythms using retention periods of 0.5 and 12.5s we also found differences between the two rhythms in memory processing (Klyn et al., submitted). Responses were more accurate to instrumental than to vocal rhythms, and accuracy differences between 'same' and 'different' decisions were different for the two rhythm types. For 'different decisions, musicians were faster than non-musicians on instrumental but slower on vocal rhythms, suggesting that musicians may make use of different stimulus representations (e.g. perceptual categorizations, or feature extractions and conceptualizations) than the non-musicians. The current study extends this research and examines the effect of concurrent motor tasks on rhythm discrimination (experiment 1) and on rhythm reproduction (experiment 2).

For both experiments we use the following experimental factors (capitalized): 'TASK' indicates the rhythm pattern to which responses are made, and has two levels, voice and clapstick; 'DISTRACTOR' refers to the concurrent motor task, and has three levels, control (no action), sub-vocal articulation and finger tapping; 'CONDITION' refers to whether the two rhythms that are compared are the 'same' or 'different'; 'STATUS' refers to musical training and experience, and has two levels, 'musician' and 'non-musician'.

Experiment 1 (Decision)

A possible explanation for differential processing of vocal versus instrumental rhythms in memorization tasks could be the existence of different memory maintenance mechanisms. For example, the phonological loop of Baddeley and Hitch's (1994) working memory model is considered to provide rehearsal and to improve retention of verbal information and consists of a phonological store and an articulatory loop. Testing the involvement of the articulatory loop in rhythm memory with a reproduction experiment, Saito and Ishio (1998) found more significant degradation with concurrent sub-vocal articulation than with concurrent finger-tapping or control. Concurrent sub-vocal articulation has been shown to decrement performance of verbal memory tasks, and is often compared to concurrent finger-tapping - the two tasks are assumed to require roughly equivalent attention and motor activation. However, because of the differences in vocal and non-vocal rhythm memory processing, we might expect to see differential effects from these concurrent motor tasks. Specifically, we would expect that finger-tapping would degrade clapstick rhythm memory more than vocal rhythm memory, and that sub-vocal articulation would degrade vocal rhythm memory more than clapstick rhythm memory. Experiment 1 was designed to test these hypotheses.

1.1. Stimuli

The stimuli for the experiments were created from field recordings of the Dyirbal from Queensland, Australia (Dixon & Koch,1996) in order to use real-world stimuli and to avoid possible confounds due to semantic processing and musical familiarity (none of our participants knew Dyirbal language or music). The stimuli that were created (mean length 1.73s, range 1.5-2s) consisted of male voice rhythms and clapstick to assure identical stimulus input for the main tasks, the comparison of two subsequent vocal or clapstick rhythm patterns. By using identical stimuli and varying the task, any difference in results can be attributed to differences in task dependent processing rather than stimuli differences. For each parent stimulus we created three variant stimuli by modifying one inter-beat interval or adding or removing one event (Hung, 2011): one with voice rhythm changed but clapstick the same, one with voice rhythm the same but clapstick rhythm changed, and one with a change to both the clapstick and the voice rhythm. The stimuli were then paired either with themselves or one of their variants, resulting in a set of 37 stimulus pairs.

For trial presentation each stimulus was followed by a silent retention interval of 12.5s and the paired second stimulus. The trials were split into two groups, each group combined with one of the three DISTRACTOR tasks (control=no motor action, finger-tapping, or repeated sub-vocal articulation of the syllable 'the'), and the resulting six blocks were presented twice, once with the instruction to respond to the voice rhythm, and once to respond to the clapstick rhythm.

1.2. Subjects

10 non-musicians (5 female; avg. age=23 years, min. 19, max. 34) and 13 musicians (8 female; avg. age=25 years, min. 21, max. 34) participated in the experiment (subjects were classified as musicians if they satisfied all of three criteria: self-identification as a musician, at least 5 years of formal musical training and current active musical engagement).

1.3. Procedure

Subjects were introduced to the experiment tasks, and given a trial section to familiarize themselves with the procedure. For the experimental runs stimuli of each block were randomly presented for each subject, and the sequence of blocks and experimental tasks were balanced across subjects. For each block the experimenter informed subjects which of the three DISTRACTOR tasks (control,

finger-tapping, or sub-vocal articulation) was to be performed, subjects started the block sequence by pressing the space key on the presentation computer, and the computer screen showed the requested TASK (vocal or clapstick rhythm). Subjects were asked to make a decision as quickly and accurately as they could, by pressing the corresponding response key on the computer keyboard. Following the experiment subjects were debriefed.

1.4. Results

Reaction Time (RT)

A repeated measurement ANOVA with a between-subject factor STATUS and within-subject factors of TASK (voice/clapstick), CONDITION (same/different) and DISTRACTOR was performed on the RT data. Only the main factors of TASK ($F_{1,21}$=9.16, p=0.006) and CONDITION ($F_{1,21}$=299.35, p<0.001) were found to be significant. The mean differences between same (1.2s) and different CONDITION (2.15s) was 0.95s, and decisions on vocal rhythms (mean: 1.63s) were 0.1s faster than those on clapstick rhythms (mean: 1.73s).

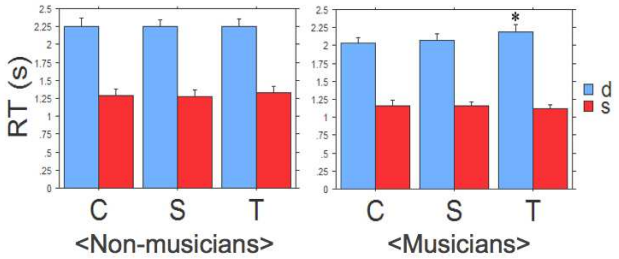


Figure 1: reaction time (RT) interaction plot for DISTRACTOR* CONDITION, split by STATUS. C=control, S=sub-vocal articulation, T=finger-tapping, d=different condition, s=same condition. Asterisk marks musicians' mean for tapping that is significantly different from their control and sub-voc response.

There was a significant 2-way interaction for DISTRACTOR and CONDITION ($F_{2,42}$=4.095, p=0.024) indicating that 'different' decisions are more affected by DISTRACTOR than 'same' decisions. A significant interaction was also found between STATUS (musician/non-musician), DISTRACTOR and CONDITION ($F_{2,42}$=4.99, p=0.011), showing that musicians are affected in their 'different' decisions by the tapping DISTRACTOR, but not non-musicians.

Accuracy

For all subjects the square root of the percentage correct decision data were subjected to an arcsin transformation and a repeated measure ANOVA with factors STATUS, TASK, DISTRACTOR and CONDITION was performed. The two main factors DISTRACTOR and TASK turned out to be significant. The factor DISTRACTOR ($F_{2,42}$=32.48, p<0.001) showed that control (77.3%) was the most accurate and tapping (67.5%) was slightly more accurate than sub-vocal (65.8%). A post-hoc analysis using Bonferroni correction showed that differences between control and tapping, and between control and sub-vocal were both significant (p<0.001), but there was no significant difference between tapping and sub-vocal (p=0.49). Subjects were less accurate when asked to complete either one of the two concurrent motor tasks than during the control, but tapping and sub-vocal affected their performance to the same degree. TASK ($F_{1,21}$=28.48, p <0.001) indicated significantly higher accuracy for the clapstick (77%) than for voice (63.5%).

A significant interaction between TASK and CONDITION ($F_{1,21}$=13.5, p=0.001) showed that while accuracy rates for clapstick rhythms were the same for 'same' and 'different' decisions (both 80%), those for voice rhythms were higher than for 'same' (70%) than for 'different' decisions (60%). Finally, the interaction between DISTRACTOR and CONDITION was also significant ($F_{2,42}$=8.445, p<0.001), showing that 'different' decisions were significantly stronger affected by the two DISTRACTOR tasks (-20% compared to control) than 'same' decisions (-10% compared to control).

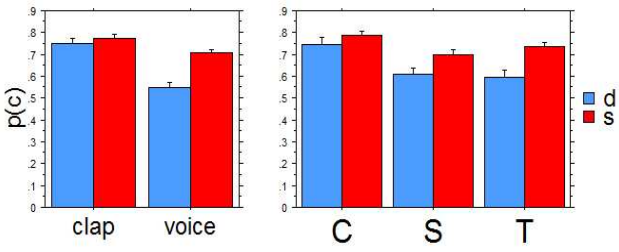


Figure 2: accuracy interaction plot for TASK*CONDITION (left) and for DISTRACTOR*CONDITION. p(c)= percentage correct, C=control, S=sub-vocal articulation, T=finger-tapping, voice=vocal rhythms, clap=clapstick rhythms, d=different condition, s=same condition.

1.5. Discussion

As expected, the additional cognitive load of the concurrent motor tasks lead to reduced accuracy in same/different decisions (control: 77.3%, tapping: 67.5%, sub-vocal: 65.8%). However, there was no significant difference between tapping and sub-vocal distractors. This result is incompatible with the hypothesis that the articulatory loop is involved in differential memorization of vocal versus instrumental rhythms. It also contrasts with Saito and Ishio's (1998) results, which showed that a concurrent sub-vocal articulation task more significantly degraded a rhythmic reproduction task than did a concurrent finger-tapping task. A possible explanation for our results could be the experimental task: subjects in our study were asked to make same/different decisions while in Saito's (1998) study they were asked to perform a motor-response task by actually reproducing the rhythms. The following experiment will test this possibility.

Although the hypothesis that the articulatory loop is generally involved in memorization of rhythms was not supported, results are consistent with those of our previous memorization experiment (Klyn et al., submitted). As expected, clapstick decisions (77%) were more accurate than the voice decisions (63.5%) overall. However, compared with the control, RT was significantly faster for voice than for clapstick rhythms for the two distractor tasks (0.095s for tap and 0.15s for sub-vocal), suggesting that for dual rhythm stimuli vocal information is processed faster if cognitive resources are constrained by simultaneous distractor tasks. Further, though the accuracy difference between musicians and non-musicians did not reach significance, non-musicians (66%) are slightly more affected by the concurrent distractor tasks than musicians (73%).

'Same' decisions are made faster and more accurately (1.2s, 75%) than 'different' decisions (2.15s, 65.5%). This is in line with results from another rhythm memory experiment (Klyn et al., submitted) and indicates that different processes and representations may be involved in same/different decisions. Additionally, there was a significant interaction between CONDITION and DISTRACTOR for both RT and accuracy. The accuracy difference between same and different decision for the two concurrent motor tasks (14%) was more than double the difference for the control task (ca. 6%), indicating that 'different' decisions are more strongly affected by the distractor tasks than 'same' decisions. This interaction is a further indication that two different processes seem to be involved in same/different decision (Keuss, 1977; Markman & Gentner, 2005): one fast, holistic, and the other slow and analytic. The slow one is cognitively more demanding (longer RT) and therefore more affected by the additional cognitive load of the distractor task.

2. EXPERIMENT 2 (REPRODUCTION)

The first experiment provided further evidence for the differential processing of vocal versus instrumental rhythms. However, the introduction of concurrent finger-tapping or sub-vocal distractor tasks did not show a significant effect on these two tasks, suggesting that the articulatory loop does not play a role in either rhythm decision tasks or in the differential processing of vocal and instrumental rhythms. Our second experiment, in which a rhythm reproduction task is substituted for the same/different decision task, will test the idea that the articulatory loop may only get recruited or be involved in rhythm reproduction tasks.

2.1. Stimuli

24 stimuli were selected from the material used in experiment 1. Each trial was setup as follows: a stimulus was followed by a silent delay interval of 12.5 s and ended with a short 1kHz beep sound as a signal for the subjects to start tapping the rhythm they had just heard. The 24 trials were split into two groups, each group combined with one of the three DISTRACTOR tasks, and the resulting blocks were presented twice, once with the instruction to tap the voice rhythm, and once to tap the clapstick rhythm.

2.2. Subjects

15 non-musicians (8 female; avg. age=24.4 years, min. 19, max. 34) and 11 musicians (5 female; avg. age=27.27 years, min. 20, max. 35) participated in this experiment ('musicians' again satisfied all of three criteria used for the previous experiment).

2.3. Procedure and equipment

Subjects were introduced to the experiment tasks, and given a trial section to familiarize themselves with the procedure. For the experimental runs stimuli of each block were randomly presented for each subject, and the sequence of blocks and experimental tasks were balanced across subjects. For each block the experimenter informed subjects which DISTRACTOR task was to be performed, subjects started the block sequence by pressing the space key on the presentation computer, and the computer screen showed the TASK (vocal or clapstick rhythm). Following the beep at the end of the delay period subjects reproduced the memorized rhythm and initiated the next trial by pressing the space key again. Subjects tapped the response rhythms with a small metal rod on a wooden tablet and their responses were recorded by the presentation software (DMDX).

2.4. Data Analysis

Responses were classified as correct if the number of response and stimulus events was identical and if the ratios of the response intervals deviated less than 20% from those of the stimulus intervals. Percentage values for the correct responses were calculated for each subject and the square root of these % values were then subjected to an arcsin transformation.

2.5. Results

A repeated measures ANOVA with between STATUS and within factors TASK and DISTRACTOR was performed. All three main factors were significant. STATUS ($F_{1,25}$=18.44, p=0.0002) showed that musicians were generally more accurate (70%) than non-musicians (46%); TASK ($F_{1,25}$=93.98, p<0.0001) indicates that there are more correct responses for clap (78%) than for vocal rhythms (38%); DISTRACTOR was also highly significant ($F_{2,25}$=17.73, p<0.0001) with mean accuracies of 63% for control, 53% for sub-vocal, and 58% for tapping; post-hoc analysis using the Bonferroni correction showed differences between the three factor levels to be

significant (control/sub-vocal : p<0.0001; control/tapping: p=0.0069; sub-vocal/tapping: p=0.0007). The difference between musicians' and non-musicians' clapstick responses under sub-vocal and finger-tapping distractor (s. fig.3) did not turn out to be significant in this study.

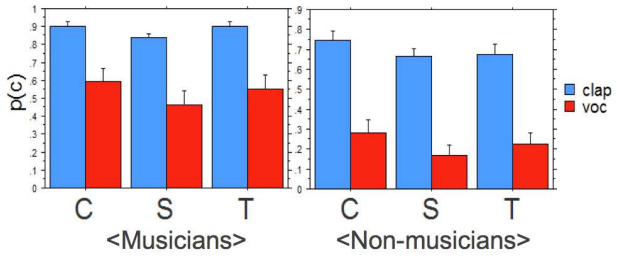


Figure 3: mean response accuracy interaction plot for DISTRACTOR and TASK, split by STATUS. p(c)= percentage correct, C=control, S=sub-vocal articulation, T=finger-tapping, voc=vocal rhythms, clap=clapstick rhythms.

2.6. Discussion

This experiment shows that a rhythm reproduction task is significantly more negatively affected by a concurrent sub-vocal articulation than by a finger-tapping task. The larger effect of the sub-vocal articulation task can be attributed to an involvement of the articulatory loop in short-term memorization, supports our initial hypothesis about such an involvement in a delayed rhythm reproduction task, and is in agreement with the findings of Saito and Ishio (1998), who report effect sizes similar to ours. However, together with the results from our experiment 1, we arrive at a different interpretation from these authors. It seems that the articulatory loop is not generally involved in rhythm memory: if rhythms are kept in memory for non-motor responses (comparison and decision) a concurrent sub-vocal articulation was not found to have a stronger effect than a concurrent tapping task (experiment 1). It is only if rhythms are memorized for subsequent motor action in a reproduction task that the articulatory loop seems to be involved. A possible explanation could be that rhythm memorization for action requires different representations and/or different processes than memorization for a decision task. Whether these different representations or processes exist simultaneously, being recruited on demand, or whether they are only created in response to the specific experimental task, remains to be addressed in further research. Earlier studies on verbal short-term memory (Morton, 1970; Levy, 1971), however, suggest coexisting dual acoustic and articulatory encodings for short term memory of verbal material, and seem to point towards the first possibility.

Similar to the decision task, musicians (70%) performed better than non-musicians (46%) in the reproduction task. The difference between them is much larger than in the decision experiment (musician: 73%, non-musician: 66%) and suggests that rhythm memory for action in non-musicians is more affected by concurrent DISTRACTOR tasks than in musicians. Furthermore, responses to clapstick rhythms (78%) are better than responses to vocal rhythms (38%), and the difference between them is larger than in the decision experiment (clapstick: 76%, vocal: 63%). Notably, there is a clear experimental task effect: clapstick rhythms are no more affected in the reproduction than in the decision task, whereas the accuracy for voice rhythms is considerably reduced. In addition, decisions on voice rhythms are more affected by the sub-vocal DISTRACTOR than clap rhythms. These results are further support for our hypothesis that clapstick and vocal rhythms have different representations in memory, and suggest that the representation of the clapstick rhythms for reproduction tasks and their representation for decision tasks may not be very different from each other, in contrast to the representations of vocal rhythms, which more strongly rely on articulatory encoding in the reproduction task.

3. CONCLUSION

The current study demonstrates that involvement of the articulatory loop is task dependent and not a general feature of rhythm memory. Hence it contributes to recent research showing that memorization of verbal (or verbalizable) and non-verbal material seems to involve, at least partly, different processes and encodings (Mercer & McKeown, 2014; Klyn et al., submitted). It also advances our understanding of auditory rhythm processing, showing that it is a process that is more complex than previously thought. Not only does rhythm processing include an important perceptual component, but there are also indications of multiple, coexisting forms of encodings that are utilized, depending on the processing stages in question as well as the experience and training of the subjects. The differences in processing between vocal and instrumental rhythms that emerge from this study offer a promising new perspective for musical rhythm research as well as for our understanding of the relationship between vocal and instrumental music.

4. REFERENCES

Baddeley, A.D., & Hitch, G. J. (1994). Developments in the concept of working memory. Neuropsychology, 8, *485-493.*

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature, 403, *309-312.*

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and non-speech sounds. Journal of Experimental Psychology: Human Perception and Performance, *32:97-103*

Dixon & Koch. (1996) Dyirbal song poetry traditional songs of an Australian rainforest people. [CD]. Mascot, N.S.W :: Larrikin.

Hung, T. H. (2011). One music? Two musics? How many musics? Cognitive ethnomusicological, behavioral, and fMRI study on vocal and instrumental rhythm processing (Doctoral dissertation, Ohio State University).

Keuss, P.I.G. (1977). Processing of geometrical dimensions in a binary classification task: Evidence for a dual process model. Perception & Psychophysics, *21:371-376.*

Klyn, N.A.M, Cheong, Y.J., Allen, E.T., and Will, U. (submitted) Differential short-term memorization for vocal and instrumental rhythms.
Levy, B.A. (1971) Role of Articulation in Auditory and Visual Short-Term Memory. Journal of Verbal Learning and Verbal Behavior, 10, *123-132.*
Markman, A.B., & Gentner, D. (2005). Nonintentional similarity processing. In R. Hassin, J.A. Bargh, & J.S. Uleman (Ed.), The New Unconscious. *(pp. 107-137) New York: Oxford University Press.*
Mercer, T., & McKeown, D. (2014). Decay uncovered in nonverbal short-term memory. Psychonomic Bulletin & Review, 21*(1), 128–135*
Morton, J. (1970) A functional model of memory. In D. A. Norman (Ed.), Models of human memory. *New York: Academic Press, 1970*
Saito, S., & Ishio, A. (1998). Rhythmic information in working memory: Effects of concurrent articulation on reproduction of rhythms. Japanese Psychological Research, 40, *10–18.*
Vouloumanos, A., Kiehl, K.A., Werker, J.F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. Cognitive neuroscience, 13, *944-1005.*
Zatorre, R.J. et al. (2002) Structure and function of auditory cortex: music and speech. Trends in Cognitive Science, 6, *37-45.*

EFFECTS OF REPETITION ON THE PERCEPTION OF MUSICAL TEXTURE

Cecilia Taher[1], René Rusch[1], Stephen Mcadams[1]
[1]McGill University, Cirmmt, Canada
Time: 17:30

1. Background: Repetition is an essential aspect of tonal music that all listeners can identify. Classical models of musical form (e.g., sonata, binary, ternary form) are practically inconceivable without the idea of repetition, and empirical studies have shown that repeated listening affects the perception of musical form (Margulis, 2012).

2. Aims: With the goal of illuminating the impact of repetition on music perception and its relevance to music theory, this paper presents an empirical investigation of the effects of structural repetition on the perception of musical texture.

3. Methods: Participants heard 18th-century-style keyboard counterpoint excerpts, composed of a repetitive and a nonrepetitive part, continuously rating the relative prominence of the two voices. To minimize potential confounding effects such as the pitch content and rhythmic profile of the two parts, 36 musical examples were especially composed. The experimental design consisted of four factors: registral position of the repetitive part (repetitive part positioned either in the lower or in the upper voice), temporal position (first or second half of each musical example), relative speed of the repetitive part with respect to the nonrepetitive part (faster than, equal to, or slower than), and length of the passage being repeated (2 or 4 measures long).

Results & Conclusions: The results indicate a shift of attention from repetitive to nonrepetitive voices that appears to be regulated by exposure to immediate repetitions of a short musical fragment. In addition, when the repetitive part is in the upper voice, it is perceived as more prominent than the non-repetitive part when it is faster than the latter, and less prominent when it is slower, whereas the opposite is the case when the repetitive part is in the lower voice. Musical repetition seems to have the special power to call the listeners' attention at first. However, as the repeated part does not change, becoming less attractive, the listeners' attention moves to other, more interesting and novel parts. In this way, repetitive music appears to dynamically shape the listeners' perception of musical texture by affecting the relative perceived importance of simultaneous musical parts (layers).

Session 6

[6A] Cognitive Neuroscience of Music 2

SEOUL 08:30-10:30 Thursday 7 Aug 2014

MUSIC INDUCES MINDFULNESS STATE -- THE SUJECHEON EFFECT

Yuh-Wen Wang1, Shan-Chuan Teng[1], Yunn-Wen Lien[2], Yun-An Huang[3], Jyh-Horng Chen[3]
[1]Graduate Institute of Musicology, National Taiwan University, Taiwan, [2]Department of Psychology, National Taiwan University, Taiwan, [3]Department of Electrical Engineering, National Taiwan University, Taiwan
Time: 08:30

ABSTRACT

The power of music over human psychological and physical states has been emphasized in East Asian classical texts. For example, two of the most important Chinese treatises on music—Yueji and Yueshu—both emphasize music's effect on enhancing people's awareness of the present moment, as well as helping them be more peaceful—an effect similar to what has been found in meditation practitioners. Yet such an effect remains untested. We thus examined whether listening to certain kinds of East Asian ancient music could change the brain temporarily in a way similar to mindfulness practices. In particular, we hypothesized that listening to a piece of Korean traditional court music, Sujecheon, could increase functional connectivity among the default mode network (DMN), a brain circuit known to be related to mind wandering, and the dorsal anterior cingulate as well as the dorsolateral prefrontal cortices (regions involving top-down control). Such enhancement was found to be one of the critical differences between long-term meditators and novices, indicating a better control of mind wandering. An fMRI experiment assessing resting states is reported, in which participants were randomly assigned to listening to either Sujecheon or a shuffled version of it, serving as a control group. The results confirm our prediction showing that the functional connectivity among DMN, anterior cingulate, and dorsolateral prefrontal cortices significantly increased during the resting state after listening to Sujecheon. In contrast, no such change was found for the control group. Our findings, for the first time, provide brain imaging data supporting the traditional East Asian view about music.

1. INTRODUCTION

The power of music over human psychological and physical states has been emphasized in East Asian classical texts. According to those texts, classical East Asian thinking appreciates that music can help people upgrade their consciousness state and achieve peacefulness (Wang, 2009, p. 203; Wang 2002; 2004a; 2012). For example, two of the most important Chinese treatises on music—Yueji and Yueshu (both written before the first century B.C.E.), both state that "When the music has full course, one becomes aware and clear with perceptive ears and eyes, and one's circulation of blood and qi becomes harmonious and peaceful." Similarly, a Korean classic text—Annals of King Sejong (written in C.E. 1454—states that the purpose of listening to ideal music is "to cultivate human nature to the loftiness of sainthood, blending the spirit and man into one, to create a universe where heaven and earth are one in accord…" (Hwang, 1985). These statements all indicate that music could enhance people's awareness of the present moment, as well as help them be more peaceful— an effect often associated with mindfulness practice. As reported, mindfulness practitioners are more likely to feel happiness and peacefulness compared to novices (e.g., Chan & Woollacott, 2007; Jha, Stanley, Kiyonaga, Wong, & Gelfand, 2010). In addition, practitioners' attentional control is improved through reducing mind wandering at rest or on task (e.g., Morrison, Goolsarran, Rogers, & Jha, 2013; Mrazek, Franklin, Phillips, Baird, & Schooler, 2013), which is also known to be associated with a sense of happiness or psychological well-being (Killingsworth & Gilbert, 2010). There is, however, no empirical study investigating such a mindfulness-like effect of music.

In the current study, we therefore aim to investigate whether listening to a piece of ancient Korean court music, Sujecheon, could produce a functional change in the brain similar to meditation practice, particularly on top-down control over mind wandering. In the following, we will first explain why Sujecheon was chosen as our target piece, followed by a short review of the meditation effect on the default mode network (DMN), the neural network that involves mind-wandering (Mason et al., 2007; Qin & Northoff, 2011; Schneider et al., 2008), and our hypothesis derived from it. Finally, an fMRI experiment assessing resting states is reported.

Sujecheon, as one of Korea's most important traditional pieces, has been intensively studied because of its cultural importance (Hwang, 1985; Lee, 1973; Wang 2004b). Different from the studies on musical analysis and historical origins of Sujecheon that most researchers focus on, our previous pilot studies showed that this piece of music might have some effects on altering listeners' consciousness state. In one of our pilot studies, 73% of 99 undergraduate students, who were not familiar with Sujecheon, were induced into mental states similar to mindfulness after listening to the entire piece of Sujecheon for 17 minutes. They reported peacefulness or better concentration after listening. In addition, three experienced meditators reported, after their first-time listening to Sujecheon, having physical and psychological reactions that they often experienced after meditation, including experiencing of a high-energy state and warmth in the back, with fluent qi circulation to sooth discomforts, as well as a sense of clearness. These subjective reports indicated that Sujecheon may be one of the pieces close to the kind suggested in the classical treaties of music, through which a mindfulness state is likely to be elicited.

In addition, according to Wang (2002), two features of Sujecheon might help listeners to be more likely to stay mindful. First, the overall musical process contains a pattern—identified as "relaxing breath tone"—which resembles the dynamic and rhythmic pattern of relaxing deep breathing—a pattern often included in meditation and qigong practices which emphasize belly breathing. Second, while there are many sectional repetitions in Sujecheon, the ever-changing rhythmic and ornamental details, together with the extremely slow tempo, encourages the listener to simply focus on the present moment—one of the main goals in mindfulness practice.

To test our hypothesis, an fMRI experiment assessing resting states was conducted to examine the change of functional connectivity regarding listeners' default mode network (DMN) during the resting state before and after listening to Sujecheon. As mentioned above, DMN refers to the neural network that involves mind-wandering, which includes dmPFC, vmPFC, posterior cingulate cortex, inferior parietal lobule, and hippocampal formation. Recently, Brewer et al. (2011) found that meditation experts had stronger functional connectivity between DMN and self-control regions (dACC, dlPFC) than did novices, both during meditation and resting periods. According to them, it indicates that meditation experts have better top-control over mind wandering and can be more "present-focusing" than novices, which is consistent with the experts'

subjective reports of reduced mind-wandering. Based on Brewer et al.'s finding, we thus hypothesized that listening to Sujecheon would increase functional connectivity between DMN and the self-control regions (especially dACC, dlPFC).

In order to control the tempo, pitches, instrumental timbre, musical style, and listening duration of music, a shuffled version of Sujecheon (will be explained in Method section) was chosen as the music piece to listen to for the control group. The shuffled version also disrupts the relaxing breath tone believed to mimic relaxing deep breathing as mentioned above. Preference ratings for the target music in each group were conducted before the experiment and showed no difference across groups. Participants who showed a strong dislike for Sujecheon were screened out before being randomly assigned to two groups, as past studies showed that enhanced post-listening performance only followed the preferred listening stimuli (e.g., Nantais & Schellenberg, 1999; Cassity, Henley, & Markley, 2007). Those who were strongly annoyed by MRI noise while listening to music were also screened out.

2. METHOD

2.1. Participants

Fifty-one healthy adults in the Taipei metropolitan area aged from 20 to 25 were recruited via Internet. The participants were randomly assigned to either the Sujecheon group (25 participants, 11 males) or the shuffled-Sujecheon group (26 participants, 14 males). Informed consent was collected and the study was approved by the Institutional Review Board at the National Taiwan University.

2.2. Design and Materials

This study adapted a single-factor (listening to *Sujecheon* or a shuffled version of *Sujecheon*), between-subject design. Participants in the *Sujecheon* group listened to a 17-minute version of Korean court music *Sujecheon* (from Selections of Korean Traditional Music, Series 1, produced by Korean Traditional Performing Arts Centre, 1988); whereas those in the shuffled group listened to a version produced by randomly reordering the segments cut in every 3.1 seconds from the same recording.

2.3. Procedures

The experiment consisted of three phases. In the 5-minute-long pre-listening resting period participants were instructed to close their eyes and not think about anything as much as they could. A 17-minute-long listening period immediately followed, during which all participants were instructed to listen to music played through earphones. Finally, during the post-listening resting period, participants were again instructed to do the same as during the first phase for 5 minutes. Participants were scanned during the entire pre- and post-listening rest periods.

After MRI scanning, participants were asked to rate their preference of the music piece they heard (5-point scale, 1 = dislike very much, 5 = like very much), the extent they were involved in listening to the music (5-point scale, 1 = not involve very much, 5 = involve very much) and the extent to which they were familiar with the music (5-point scale, 1 = not at all, 5 = very much).

2.4. MRI Scanning and Equipment

All of the images were conducted on a Bruker (Ettlingen, Germany) 3T Medspec system. For the functional images, a gradient echo-planar imaging (EPI) sequence with BOLD (blood oxygenation level dependent) was used. Functional images were collected parallel to AC-PC plane with whole brain EPI acquisition from bottom to top in an interleaved sequence. The following parameters were used: TR = 3000 ms, TE = 30 ms, FOV = 24 x 24 cm, thickness = 3.75, number of slices = 35. In the listening period, a MR compatible ultra-fidelity headset (Resonance Technology Inc, Los Angeles, U.S.A.) was used to present the music.

2.5. MRI data analysis

Data analysis was performed using SPM8 (Statistical Parametric Mapping, UCL, London, U.K.) and REST toolkit (RESting-state fMRI data analysis Toolkit, Zhejiang, China). The functional images were realigned to the first volume and normalize to the MNI template. Detrend and bandpass filter (0.01 Hz–0.1 Hz) was implemented to filter out the respiratory and cardiac signal. Signal changes in the CSF and white matter was excluded to avoid the artifacts from non-BOLD signal fluctuations. The spatial smooth kernel is 8 × 8 × 8 mm. The seed region of default mode network (DMN) is the posterior cingulate cortex (PCC) located at center [0, -52, 30] with 5 mm radius in MNI coordinate. All the reported results satisfied the threshold $p < .05$ FDR corrected.

3. RESULTS

After preprocessing, 13 participants were excluded from further analysis. Twelve were due to large motion artifacts, and the remaining one was excluded due to not showing functional connectivity of any neural circuit. The following analysis were based on the remaining 38 participants, 18 in the *Sujecheon* group (Mean age = 21.8 years old, SD = 1.7) and 20 in the shuffled group (Mean age = 22.2 years old, SD = 1.6). There was no significant difference in age across conditions ($p > .4$). There was also no significant difference in preference, involvement in music listening, or familiarity with music across conditions, $ps > .2$.

Table 1: functional connectivity with posterior cingulate cortex seed region for the *Sujecheon* group (post-listening resting period > pre-listening resting period)

brain area	T	cluster	x	y	z
R parahippocampal gyrus	8.53	1844	22	-80	0
L parahippocampal gyrus	7.25	983	-24	-50	-8
R dACC	6.6	51	14	42	2
R uncus	5.88	151	20	-2	-34
R dlPFC*	5.66	57	16	62	32
L dlPFC*	5.34	82	-14	60	6

Notes. Results are P < .05 FDR-corrected.
*indicates regions of a priori interest, reported at uncorrected P < .001.

To test our hypothesis, we compared the functional connectivity of DMN in the resting states before and after the listening period. For the participants listening to *Sujecheon*, we found functional connectivity between PCC, used as the seed region, and right dACC, parahippocampal gyri (PHG), as well as right nucus were significantly increased (Table 1 and Fig. 1). In addition, we also found increased connectivity between PCC and dlPFC in the threshold $p < .001$ uncorrected (Table 1 and Fig. 1). In contrast, there was no significant increased functional connectivity between PCC and other brain regions after listening to the shuffled-version of *Sujecheon*.

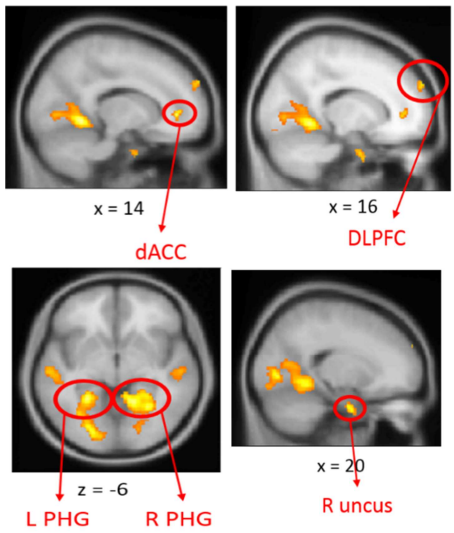


Fig. 1: Functional connectivity with posterior cingulate cortex seed region (post-listening resting period > pre-listening resting period) for the *Sujecheon* group.

4. DISCUSSION

As predicted, a stronger coupling between DMN and the self-control regions (dACC, dlPFC) was found after listening to *Sujecheon* but not after the shuffled piece, a difference similar to what has been found between long-term meditation practitioners and novices (Brewer et al.,2011). In addition, functional connectivity within DMN areas were also increased. The difference found between groups cannot be attributed to participants' preference or attention to the music heard since their ratings of preference and involvement were the same across groups. Our findings not only support our hypothesis that listening to certain traditional East Asian music is likely to induce a more present-focused resting state, but also confirms the traditional East Asian view about the function of ideal music for the first time.

It is noteworthy that such changes of functional connectivity were found after, but not during, music listening. In other words, the present results revealed a sustained effect of *Sujecheon* rather than an instant one. As the listening process goes by, the co-activation among those brain areas might gradually increase and reach a peak at the end of the piece, and last for a short period after listening. If so, there should be a time-series change of the strength of functional connectivity between those brain regions during listening, which needs further investigation. In addition, since it is the "sustained" effect of music listening, it would also be interesting to know how long such an effect would last.

Although the present result is intriguing, we still have to be cautious about its implications, since the underlying physiological mechanism of increased functional connectivity remains unclear. Behavioral evidence is still needed to further support our hypothesis.

Moreover, what contributes to the meditation-like effect of *Sujecheon* still requires further investigation. Although the null effect of the shuffled-version might come from a lack of the relaxing breath tone (RBT) pattern we manipulated, future studies that directly reveal the relationship between RBT and the listeners' respiratory patterns are still required.

Although we try to link the effect of traditional East Asian music to mindfulness practices, it is worth noting that we are not arguing that the *Sujecheon* effect is as strong and stable as long-term meditation practice. It is clear that the former is more cue-driven and the latter is more actively involved. We tend to regard this kind of music as a temporary facilitator to mindfulness. It would be interesting to know whether listening to *Sujecheon* would also lead to any enhancement at the behavior level as mindfulness practice usually does, such as improving emotional regulation or cognitive functions. It is also practically useful to examine whether listening to *Sujecheon* is helpful for the novice to reach mindfulness state during meditation practice.

5. REFERENCES

Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y. Y., Weber, J., & Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. Proceedings of the National Academy of Sciences, *108(50), 20254-20259.*

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. Journal of personality and social psychology, *84(4), 822-848.*

Cassity, H. D., Henley, T. B., & Markley, R. P. (2007). The Mozart Effect: Musical Phenomenon or Musical Preference? A More Ecologically Valid Reconsideration. Journal of instructional Psychology, *34(1).*

Chambers, R., Lo, B. C. Y., & Allen, N. B. (2008). The impact of intensive mindfulness training on attentional control, cognitive style, and affect. Cognitive Therapy and Research, *32(3), 303-322.*

Chan, D., & Woollacott, M. (2007). Effects of level of meditation experience on attentional focus: Is the efficiency of executive or orientation networks improved? The Journal of Alternative and Complementary Medicine, *13(6), 651-658.*

Hwang, B. K. (1985). Some notes on Korean music and aspects of its aesthetics. WORLD OF MUSIC, 27*(2), 32-48.*

Jha, A. P., Stanley, E. A., Kiyonaga, A., Wong, L., & Gelfand, L. (2010). Examining the protective effects of mindfulness training on working memory capacity and affective experience. Emotion, 10*(1), 54-64.*

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. Science, *330(6006), 932-932.*

Lee, S.C. (1973). Samhyon Yukkak. Survey of Korean Arts: Traditional Music.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. Science, *315(5810), 393-395.*

Morrison, A. B., Goolsarran, M., Rogers, S. L., & Jha, A. P. (2013). Taming a wandering attention: Short-form mindfulness training in student cohorts. Frontiers in Human Neuroscience, *7.*

Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance while reducing mind wandering. Psychological Science, *24(5), 776-781. doi: 10.1177/0956797612459659*

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. Psychological Science, *10(4), 370-373.*

Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? Neuroimage, *57(3), 1221-1233.*

Schneider, F., Bermpohl, F., Heinzel, A., Rotte, M., Walter, M., Tempelmann, C., . . . Northoff, G. (2008). The resting brain and our self: self-relatedness modulates resting state neural activity in cortical midline structures. Neuroscience, *157(1), 120-131.*

Wang, G. (2009). Zhong xi yinyue zhi yi tong (The similarities and differences between Chinese and Western music). In W. Feng & Y. Yu (Eds.), Selections of Guangqi Wang's writings on music. *Beijing: People's Music Publishing House.*

Wang, Y. (2002). Music, Mind-Body States, and Ethos—Experiences of the Korean Court Music Sujecheon. Essays of The 2001 Conference, Society for Ethnomusicology, Republic of China, *pp. 62-77.*

Wang, Y. (2004a). The ethical power of music: Ancient Greek and Chinese thoughts. The Journal of Aesthetic Education, 38*(1), 89-104.*

Wang, Y. (2004b). Issues of Tension in Some Traditional Asian Music. *Paper presented at the 37th World Conference, International Council for Traditional Music, Fujian, Quanzhou.*

Wang, Y. (2012). Music as Moral Discipline: Yue Ji and the Chinese Literati Tradition of Guqin. In W. Bowman & A. L. Frega (Eds.), The Oxford Handbook of Music Education Philosophy*: Oxford University Press.*

MUSICAL CONTEXT CHANGES NEURAL RESPONSES TO TONIC TRIAD

Jeong Mi Park[1], Suk Won Yi[1,3,7], Chun Kee Chung[2,4,5], June Sic Kim[2], Kyung Myun Lee[6], Jaeho Seol[7]
[1]Interdisciplinary Program in Musicology, Korea, Republic of, [2]Meg Center, Department of Neurosurgery, Seoul National University Hospital, Korea, Republic of, [3]Western Music Research Institute, Seoul National University, Korea, Republic of, [4]Department of Neurosurgery, Seoul National University, Korea, Republic of, [5]Interdisciplinary Program in Cognitive Science, Seoul National University, Korea, Republic of, [6]Graduate School of Convergence Science and Technology, Seoul National University, Seoul National University, Korea, Republic of, [7]6imaging Language Group, Brain Research Unit, O. V. Lounasmaa Laboratory, Aalto University School of Science, Aalto, Finland
Time: 09:00

1. Background: Previous studies have reported experienced tones enhance auditory representations around 100 to 200 ms. Those studies, however, didn't concern a contextual effect. Most studies for harmonic context have reported that harmonic processing is related to frontal or parietal cortices at a cognitive level. Given behavioral studies showing that harmonically more expected chord makes us to perform faster than others by a priming effect, we anticipated that musical context influences perception of tones.

2. Aims: The present study aimed to investigate the influence of musical context on auditory neural responses to tonic chord by using magnetoencephalography (MEG)

3. Method: To examine our hypothesis, we measured the auditory evoked fields (AEFs) for tonic chords at three different positions (the first, second, and last) among a five-chord progression. The second chord is expected by the simple repetition of the first, whereas the last by a perfect authentic cadence.

4. Results: The results showed that in the temporal regions, the P2m amplitudes were the smallest for the second, while the P2m latencies were shortest for the last. This effect was found to be greater in musicians than nonmusicians. The results show that simple repetitions are reflected in P2m amplitudes, and harmonic expectation is reflected in P2m latencies. An ERAN was found only in musicians.

5. Conclusions: Our results newly reveal that musical context influences auditory representations.

NEURONAL PROCESSES DURING IDENTIFICATION OF FAMILIAR MELODIES. INFLUENCES OF MUSICAL EXPERTISE AND SYNTACTIC STRUCTURES

Niklas Büdenbender[1], Gunter Kreutz[1]
[1]Carl von Ossietzky University Oldenburg, Department of Music, Germany
Time: 09:30

1. Background: The application of psycholinguistic principles such as cohort model and gating paradigm in melody cognition has shed some light on the mental processes contributing to the aural identification of tunes. Specifically, Dalla Bella et al. (2003) thus found that the number of tones needed to identify a played melody as either familiar or unfamiliar was influenced by musical expertise as well as structural cues, i.e. motive boundaries. Eventually, Daltrozzo et al. (2010) investigated neuronal activity in an event-related potential (ERP) study at the moment where listeners first report a feeling of knowing—specified as the familiarity emergence point (FEP). Results revealed greater amplitudes of early brain potentials at the FEP gate compared with those succeeding it. An increased fronto-central negativity was also found for highly familiar melodies, peaking around 400 msec. The authors concluded that these components might reflect conceptual processing of melodies.

2. Methods: We investigated ERPs around the FEP in cohorts of musicians (n = 18) and nonmusicians (n = 15) in a gating paradigm study. We further measured neuronal activity at the moment of definitely identifying a melody as familiar—specified as the familiarity point (FP).

3. Results: Our findings are similar to those of Daltrozzo et al. regarding early brain potentials at and around the FEP. Furthermore, musicians showed a late positive potential emerging at centro-parietal sites during the FP. This bears resemblance to a music closure positive shift (music CPS) that is usually associated with syntactic caesuras such as motive boundaries (cf. Neuhaus et al., 2006).

4. Discussion: Our results suggest that musical expertise differently affects EEG responses in early vs. late stages of melody identification processes. Contrary to non-musicians, musicians appear to anticipate mental closure of syntactic musical structures. Implications of these findings with respect to brain signatures of conceptual processes in melody recognition, as well as the necessity of a more differentiated view on familiarity as a moderating variable will be discussed.

OSCILLATORY BRAIN DYNAMICS IN THE PROCESSING OF IMPLIED HARMONY

Jung Nyo Kim[1], Ji Chul Kim[1], Edward Large[1]
[1]University of Connecticut, USA
Time: 10:00

1. Background: A large number of electrophysiological studies on harmonic perception have investigated the time course of harmonic processing by analyzing event-related potentials (ERPs), but little is known about dynamic patterns of underlying neural oscillations. However, a few studies have shown that low frequency synchronization in delta (< 4 Hz) and theta (4 – 8 Hz) bands is related to the processing of harmony in chord progressions.

2. Aims: The present study investigates the oscillatory brain dynamics in harmonic processing by means of wavelet based time-frequency representation (TFR) analysis. In particular, we focus on implied harmony in single-voice melodies rather than explicit harmony in chord progressions.

3. Method: In a previous study (Kim, Large, & Ashley, 2013), participants heard tonal melodies of which the final implied chords were manipulated. The final target chord following I-V-I-ii-V chord progression was either expected I or unexpected IV and started either on time or one beat early. It was found that these manipulations modulated amplitudes of the ERAN (100 – 250 ms) and the N5 (400 – 600 ms) elicited by the target tones. Here, TFR analyses were conducted in time windows of the ERAN and the N5 over the prefrontal region.

4. Results: The results showed a stronger increase in delta and theta band power for the unexpected subdominant chord tones than for the expected tonic chord tones in the ERAN time window. However, increases in delta and theta in the same time window were weaker for the unexpected early chord changes than for the expected on-time changes. In the N5 time window, oscillatory activity in theta band increased for the expected chord tones and decreased for the unexpected chord tones.

5. Conclusions: These oscillatory changes in delta and theta bands support the ERP results of the previous study that suggested that the ERAN was influenced not only by the violation of harmonic expectation but also by the metric position and that the N5 only reflected the difficulty of harmonic integration caused by harmonic expectation. Moreover, the results of the present study confirm and extend the previous findings that the processing of harmony modulates oscillatory dynamics in delta and theta bands.

[6B] Multimodal Perception and Performance

KYOTO 08:30-10:30 Thursday 7 Aug 2014

SIGHT & SOUND: CROSSMODAL INTERACTIONS IN THE PERCEPTION OF MUSICAL PERFORMANCE

Jonna Vuoskoski[1], Eric Clarke[1], Charles Spence[1], Marc Thompson[2]
[1]University of Oxford, United Kingdom, [2]University of Jyväskylä, Finland
Time: 08:30

1. Background: Music is an inherently multisensory phenomenon, and the body movements and gestures of a performer can communicate a range of meaningful information to the observer of a musical performance. Although the importance of body movements and gestures has been highlighted by previous research, little is currently known about the relative significance of auditory and visual performance cues from the observers' point of view. Furthermore, the investigation of possible audiovisual interactions has been held back by the technical difficulties associated with the generation of controlled, mismatching stimuli.

2. Aims: First, the aim was to investigate the relative contributions of auditory and visual performance cues to a) observers' perception of performance expressivity, and b) the subjective emotional responses of observers. Second, the aim was to investigate potential crossmodal interactions in a) the perception of visual and auditory expressivity, and b) the perception of tempo and loudness variability.

3. Method: A total of 72 participants took part in four experiments. Motion-capture animations and audio recordings of two pianists performing a Chopin Prelude (with three different expressive intentions) were used as the stimulus material. A novel method was utilized to generate controlled, synchronized stimuli that comprised both matching and mismatching bimodal combinations of the different expressive intentions.

4. Results: The results revealed that – in the case of perceived expressivity – the effect of visual kinematic cues may be somewhat stronger than that of auditory performance cues. However, in the case of subjective emotional responses, the roles of auditory and visual kinematic performance cues seemed to be more evenly balanced. In certain performance conditions, visual cues had an effect on the ratings of auditory expressivity, and auditory cues had a small effect on the ratings of visual expressivity. The results also revealed that the perception of loudness variability was affected by visual kinematic cues.

5. Conclusions: Our findings highlight the importance of visual cues in the perception of musical performance, and provide preliminary evidence for the presence of genuine crossmodal effects in the perception of certain auditory features. Our findings also suggest that non-auditory cues may contribute more to music-induced emotions than previously thought.

SIGHT, SOUND AND SYNCHRONY: EFFECTS OF ATTENUATING AUDITORY FEEDBACK ON DUO VIOLINISTS' BEHAVIOURS IN PERFORMANCE

Robert Fulford[1], Carl Hopkins[2], Gary Seiffert[2], Jane Ginsborg[1]

[1]*Royal Northern College of Music, United Kingdom,* [2]*University of Liverpool, United Kingdom*
Time: 09:00

ABSTRACT

Hearing impairments affect not only the auditory perception of music but also performers' behaviours while playing together. This study explored the effects of systematically attenuating auditory feedback on ensemble synchrony, spontaneous looking behaviours and overall dynamic levels when playing. Four violinists without hearing impairments formed two duos and played in two acoustically-isolated rooms separated by a glass window. Each violinist was able to hear their own playing and their co-performer's playing with different combinations of auditory attenuation (0dB, -10dB, -20dB, -30dB and -40dB). Video-recordings of the violinists' looking behaviours were coded and the sound pressure levels of the violinist's playing were measured. Signed asynchronies were extracted from waveform peaks observed in the wav files. The results showed that the violinists' looking behaviour was not affected by attenuating the auditory feedback from their own playing but was affected by attenuating that from their co-performer's playing: the more the attenuation, the more often they glanced and the longer they gazed at their co-performers. The more the audio feedback from the violinists' own playing was attenuated, the louder they played; conversely, the more that from their co-performer's playing was attenuated, the more quietly they played. There were no effects on ensemble synchrony. The results suggest that it is not the level of auditory feedback *per se*, but rather the ratio of perceived loudness between the player and their co-performers, which affects communicative looking behaviours and loudness of playing.

1. INTRODUCTION

The main challenges for deaf musicians are staying in time and staying in tune with other players. According to interview data, these are addressed, first, by idiosyncratic, dynamic sensory attention to visual, physical, proprioceptive and vibrotactile information and, second, via sensory compensation, particularly with respect to visual and auditory information (Fulford, Ginsborg, & Goldbart, 2011). Observational research on non-hearing-impaired duos (violin and flute-piano) wearing standard foam ear plugs that attenuate auditory feedback (by which we mean the audio signal produced by both musicians) by ~35dB and are recommended for use by musicians to prevent noise-induced hearing loss (NIHL), suggests that mild or moderate hearing impairment does not reduce performers' ensemble synchrony or increase their looking behaviours (Fulford, Ginsborg, & Goldbart, 2012). Similar research with duos including congenitally, profoundly deaf musicians showed that they looked more at their co-performers while making music, thereby prompting reciprocal looking behaviour (Fulford & Ginsborg, 2013, 2014 in press). In extreme cases dynamic sensory attending may produce sensory compensation. For example, where both members of a flute-piano duo were profoundly deaf, their ensemble synchrony was reduced and their looking behaviours increased as they became aware of, and attempted to resolve their asynchrony. Apparently spontaneous strategies such as these are likely to have been learned over time and reflect an increased reliance on visual information in daily life generally, and particularly for the purposes of communication.

Empirical attempts to demonstrate superior visual abilities in deaf people have produced mixed results, due to heterogeneous samples and confounding variables. Longitudinal research shows that people born profoundly deaf develop different abilities at different times; visual compensation for deafness may not develop until adulthood (Rettenbach, Diller, & Sireteanu, 1999). One laboratory study found that deaf individuals "possessed greater attentional resources in the periphery [of the visual field] but less in the centre when compared to hearing individuals" (Proksch & Bavelier, 2002, p. 687). Another found 'enhanced function' effects but only in a small sample of (congenitally) deaf native signers (Bavelier, Dye, & Hauser, 2006). Bosworth and Dobkins (2002) compared signing and non-signing deaf and hearing participants' discrimination of direction-of-motion. The strong right visual field advantage shown by signers suggests that perceptual processes are related to language, processed in the left hemisphere of the brain, and the inferior and peripheral visual field advantages shown by deaf participants are attributable to auditory deprivation. These findings, and the hypothesis that sensory compensation develops over time, are supported by more recent research showing differences between profoundly deaf and hearing individuals in their retina and optic nerves at the neural level, prior to the visual cortex that is responsible for peripheral vision (Codina et al., 2011). Interactive musical performance is typical of a complex situation making high attentional demands on players and therefore requiring the employment of a range of strategies. Yet it would seem that we are predisposed to using our eyes in such situations, especially when auditory feedback is compromised. According to Morgan et al. (2011), "without an increase in attention to the auditory stimuli, visual stimuli remain prepotent" (p. 13).

1.1. Aims and research questions

This study aimed to identify the level of attenuation of auditory feedback at which ensemble synchrony is likely to be impaired, and/or looking behaviours increase in response to the demands of musical performance, by replicating Fulford et al.'s (2012) study in an acoustically-controlled environment. It also aimed to explore the potential effects of auditory attenuation on sound pressure levels produced by players ("loudness") since this had been mentioned anecdotally by both deaf and hearing participants in previous studies by the same authors. Three research questions were posed; the answers could be of use in understanding the needs of musicians with diagnosed hearing impairments resulting, for example, from NIHL. Hypotheses derive from the findings of previous research.

1. **What is the effect of attenuating auditory feedback on players' ensemble synchrony?** It was hypothesised that players ensemble synchrony would be *worse* the *more*
    a. their own feedback was attenuated
    b. their partners' feedback was attenuated
    c. their partners' feedback was attenuated in relation to that of their own.

2. **What is the effect of attenuating auditory feedback on players' looking behaviours?** It was hypothesised that players would look towards each other *more* the *more*
    a. their own feedback was attenuated
    b. their partners' feedback was attenuated
    c. their partners' feedback was attenuated in relation to that of their own.

3. **What is the effect of attenuating auditory feedback on the sound pressure levels produced by the players?** It was hypothesised that players would produce higher sound pressure levels (i.e. play *louder*)
    a. the *more* their own feedback was attenuated
    b. the *less* their partners' feedback was attenuated
    c. the *less* their partners' feedback was attenuated in relation to that of their own.

2. METHOD

2.1. Participants

Two male and two female violinists aged between 23 and 26, students on the BMus and MMus course(s) at the Royal Northern College of Music, UK, were recruited, and formed two duos. In both cases the male violinist played the first violin part.

2.2. Design

Previous studies manipulating the auditory feedback to participants and/or investigating ensemble synchrony have used MIDI outputs from electronic pianos to allow researchers to obtain and quantify temporal data (Goebl & Palmer, 2009; Keller & Appel, 2010). In order to manipulate acoustic auditory feedback in this study, each violinist played in an acoustically isolated room in the recording studio at the Royal Northern College of Music. Each room had a Brüel and Kjær Type 2231 sound level meter to record each violinist from which the AC output was sent to the mixing desk and a two-channel DAT recorder. An observation window between the two rooms allowed a direct line-of-sight between the two violinists and allowed each performer to see the head, torso and arms of their co-performer as well as their violin and bow. Auditory feedback into the ears of each violinist was relayed using in-ear headphones (TDK EB900) because this allowed control of an audio mix comprising the sound from the performer's and the co-performer's violins. However, this required the radiated sound from each performer's violin to be highly attenuated by using hearing protection to avoid the performer hearing this sound via transmission through the shell of the in-ear headphones or air gaps between these headphones and the ear canal. For this reason each player wore a combination of in-ear headphones and hearing defenders (Peltor H540A (L)) that, in combination, provided high attenuation. The mixing desk was used to alter the auditory feedback conditions and the DAT recording was used to post-process the recordings to quantify the levels of each violinist during the performances.

Preliminary acoustic tests were performed to establish: i) the baseline sound pressure level from the violin at the ear; ii) the attenuation provided by the ear defenders and/or the in-ear headphones; iii) the acoustic isolation between the two rooms; iv) the listening levels from the mixer output of the violinists' playing and, finally; v) the subjective experience of manipulating the mixer output level according to experimental conditions. These tests involved the use of a Brüel and Kjær Type 4100 Head and Torso Simulator, a Type 4157 Ear Simulator and standard audiometry using Octovation Amplitude T-Series audiometer. Whilst a full account of these acoustical measurements is not possible here, the tests were used to identify what could be achieved in terms of manipulating auditory feedback. The combined attenuating effect of wearing in-ear headphones and hearing defenders (Figure 1) meant that it was feasible to use levels of -40dB as the highest level of sound attenuation of a violinist's co-performer, and up to -30dB for the violinist's *own* playing. This ensured that the level of the radiated sound from a performer's violin was at least 10dB below the listening level at the ear. The building construction between the two rooms in the studio provided at least 43dB of attenuation at all relevant frequencies for the violin.

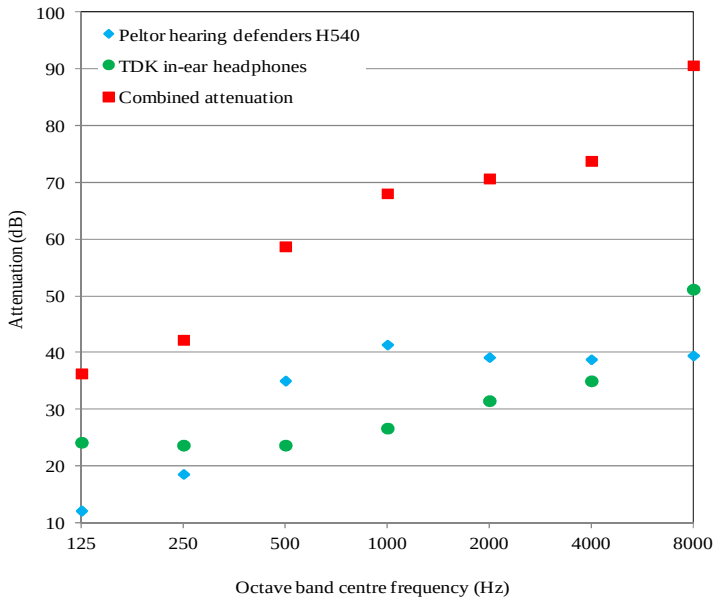


Figure 1. Attenuation provided by in-ear headphones, hearing defenders and the combination of the two.

2.3. Procedure

Each participant learned a new work composed for the purposes of the research by Emma-Ruth Richards, *Sketch* (duration: 2 minutes). The commission included 'entry markers' and tempo changes for each player individually and both players. Panasonic NV-GS280 video cameras were used to record performances. The main study used a matrix of 20 feedback conditions (Table 1) involving all possible combinations of: four conditions in which the player heard his or her own playing at a 'normal' sound level and attenuated by 10dB, 20dB and 30dB; and five

conditions in which the player heard his or her co-performer at the same four sound levels and, in addition, attenuated by 40dB, this last condition was made possible by the acoustic isolation between the two rooms. The 20 conditions were presented in random order, and each duo performed *Sketch* in full, in every condition.

Table 1. Matrix of conditions

	Player			
Co-performer	**0dB**	**-10dB**	**-20dB**	**-30dB**
0dB	1	2	3	4
-10dB	5	6	7	8
-20dB	9	10	11	12
-30dB	13	14	15	16
-40dB	17	18	19	20

2.4. Analyses

Ensemble synchrony, looking behaviours and overall dynamic level of playing were measured in all 20 conditions. The post-processing from the two sound level meters that recorded each violinist was carried out by converting the DAT recordings to wav files. These were then played using Adobe Audition into a Brüel and Kjær Type 2260 sound level meter. Calibration tones on the DAT recording allowed calibrated levels to be measured using the sound level meter for (a) the A-weighted equivalent continuous sound pressure level, L_{Aeq}, and (b) the A-weighted Fast time-weighted maximum sound pressure level, L_{AFmax}. Baseline levels were established by calculating the mean of the sound levels produced in a practice run undertaken by each duo and performance in the condition in which the auditory feedback available to both performers was unattenuated. Relative average and peak loudness levels could then be calculated in relation to the baseline. Signed asynchronies were extracted from waveform peaks visualised from the wav files in Audacity as follows: the first violin part was used as the referent. If the second violin part lagged behind the first, the signed asynchrony was negative; if it 'overtook' the first, the signed asynchrony was positive. First, the frequency and duration of partner-directed glances and gazes through the window, as observed in the video-recordings, were measured using Noldus Observer XT9. The levels of auditory feedback derived from a) the performer's own, b) his or her partner's playing and c) the difference between them were collapsed into three conditions respectively: Performer – 1) 0dB (no attenuation), 2) -10 and -20dB, and 3) -30dB; Partner – 1) 0dB, 2) -10 and -20dB, and 3) -30 and -40dB; Difference – 1) +30 to +10dB, 2) 0 and -10 dB, and 3) -20 and -40dB. Second, Spearman's correlations were calculated to identify relationships between the dependent variables and levels of auditory attenuation. Third, the effects on the dependent variables (ensemble synchrony, looking behaviours and overall dynamic level of playing) of each condition were explored using unrelated analyses of variance or Kruskal-Wallis tests, with planned contrasts (ascending groups) or Jonckheere's trend tests to confirm the relationships previously identified.

3. RESULTS

As shown in Table 2, attenuating the auditory feedback from the performer's partner was significantly associated with more frequent glances and longer gazes, and attenuating the auditory feedback derived from the performer was significantly associated with playing at a higher overall sound pressure level; no such association was found between attenuating auditory feedback and ensemble synchrony.

Table 2. Spearman's correlations between auditory feedback and measured variables

	Auditory feedback level		
	Player	Partner	Diff
Looking (frequency)	.154	-.297**	-.331**
Looking (duration in s)	.110	-.311**	-.315**
Sound pressure level in terms of L_{Aeq} (dB)	-.279***	.344***	.468***
Sound pressure level in terms of L_{AFmax} (dB)	-.251***	.188**	.320***
Signed asynchronies (ms)	-.092	-.007	.041
Unsigned asynchronies (ms)	.071	-.004	-.061

* < .05, ** < .01, *** < .001

Subsequent analyses of variance testing the effects of attenuating auditory feedback grouped as shown in Table 3 were used to answer the research questions as follows.

1. What is the effect of attenuating auditory feedback on players' ensemble synchrony? No significant effects were found on ensemble synchrony of attenuating the players' own feedback, their partners', or the players' own feedback in relation to that of their partners, so none of the hypotheses was supported.

2. What is the effect of attenuating auditory feedback on players' looking behaviours? No significant effect was found on looking behaviours of attenuating the players' own feedback so the first hypothesis was not supported. Players looked towards each other more, however, as their partner's feedback was increasingly attenuated, both in absolute terms and in relation to their own feedback, supporting the second and third hypotheses**.**

3. What is the effect of auditory feedback on players' sound pressure levels? Players' sound pressure levels (in terms of L_{Aeq} and L_{AFmax}) increased the more the feedback from their own playing was attenuated and less that of their partners' feedback was attenuated, both in absolute terms and in relation to their own feedback, supporting all three hypotheses.

Table 3. Results of ANOVAs and Kruskall Wallis tests showing the effects of auditory feedback levels on dependent variables

	Player level M (SD)			**Test statistic, (effect size)**	***p***
	0dB	**-10dB and 20dB**	**-30dB**		
Looking (frequency)	10.6 (7.5)	9.0 (5.6)	8.2 (4.9)	F (2, 81) = 0.84	.378
Looking (duration in s)	12.8 (9.0)	10.9 (6.9)	10.4 (5.2)	F (2, 81) = 0.65	.471
Sound pressure level in terms of L_{Aeq} (dB)	-0.6 (1.9)	0.2 (2.2)	0.5 (2.1)	***H* (2) = 16.23, (r^2 = .24)**	**.001**
Sound pressure level in terms of L_{AFmax} (dB)	-0.4 (2.1)	0.7 (2.6)	0.9 (2.7)	***F* (2, 249) = 5.85, (ω^2 = .04)**	**.003**
Signed asynchronies (ms)	0.0 (0.1)	0.0 (0.1)	0.0 (0.1)	F (2, 249) = 0.52	.597
Unsigned asynchronies (ms)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	H (2) = 2.88	.237
	Partner level M (SD)				
	0dB	**-10dB and -20dB**	**-30dB and -40dB**		
Looking (frequency)	7.2 (5.3)	8.2 (4.9)	11.6 (6.9)	***F*(2,81) = 8.01, (ω^2 = .08)**	**.006**
Looking (duration in s)	8.7 (5.9)	10.2 (6.2)	14.1 (8.2)	***F*(2,81) = 4.40 (ω^2 = .08)**	**.005**
Sound pressure level in terms of L_{Aeq} (dB)	1.0 (1.6)	0.4 (1.8)	-0.9 (2.3)	***H* (2) = 28.73 (r^2 = -.34)**	**.001**
Sound pressure level in terms of L_{AFmax} (dB)	1.1 (2.5)	0.7 (2.4)	-0.3 (2.6)	***H* (2) = 8.76 (r^2 = -.18)**	**.007**
Signed asynchronies (ms)	0.2 (1.0)	0.0 (0.8)	0.0 (0.1)	F (2, 249) = 0.64	.527
Unsigned asynchronies (ms)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	H (2) = 2.27	.321
	Difference level M (SD)				
	+30+10dB	**0dB and -10dB**	**-20dB and -40dB**		
Looking (frequency)	7.8 (4.9)	7.8 (5.1)	12.9 (7.0)	***F* (2, 81) = 9.80, (ω^2 = .13)**	**.002**
Looking (duration in s)	9.6 (5.7)	9.6 (6.1)	15.7 (8.5)	***F* (2, 81) = 9.81, (ω^2 = .13)**	**.002**
Sound pressure level in terms of L_{Aeq} (dB)	1.3 (1.8)	0.0 (1.6)	-1.2 (2.3)	***H* (2) = 54.75, (r^2 = -.48)**	**.001**
Sound pressure level in terms of L_{AFmax} (dB)	1.7 (2.7)	0.2 (2.1)	-0.5 (2.6)	***F* (2, 249) = 16.67, (ω^2 = .11)**	**.001**
Signed asynchronies (ms)	-0.0 (0.1)	0.0 (0.1)	0.0 (0.1)	H (2) = 1.39	.499
Unsigned asynchronies (ms)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	H (2) = 0.75	.688

4. DISCUSSION

The strongest effects found in this study were those resulting from manipulating differences between the levels of auditory feedback made available to players and their co-performers. Perceived 'self-to-other ratio' (SOR) can be an important factor in the ways musicians blend the sounds of their respective instruments in an ensemble (Keller, 2014). For example, Ternström (2003) found that, on average, choral singers prefer a SOR of +6dB (i.e. when their own feedback was at a level 6dB higher than that of their fellow singers). In the present study, looking behaviours and loudness of playing as inferred from sound pressure levels were measured, rather than preference; these were clearly affected by SORs albeit larger than those found by Ternström: +40dB and -30dB

The harder it was for the participants to hear their co-performers in relation to their own playing the more they looked towards them. In group performance, then, visual contact may well be a useful strategy for maintaining ensemble synchrony when players anticipate difficulties arising from poor auditory feedback. Visual contact began to increase, in the present study, only when SORs were equal to or more than +20dB (i.e. when the sound pressure level produced by the co-performer was four times as quiet as that produced by the player). The attenuation of auditory feedback affected players most notably in that they played louder the harder it was for them to hear themselves, and quieter the harder it was for them to hear their partners. Ensemble synchrony remained unaffected even when the sound pressure levels produced by co-performers were reduced by 40dB compared to the player's own level (NB Every reduction of 10dB is perceived as halving the loudness). It has already been shown that profoundly deaf musicians are able to stay in time by looking more towards co-performers (Fulford & Ginsborg, 2014 in press) and the present results support the idea that temporal synchrony between musicians remains intact when auditory feedback is poor, in part because players start using visual information such as the movements and gestures of their co-performers when SORs are +20dB or more.

In sum, the 'self-to-other ratio' is an important consideration for deaf or hard-of-hearing musicians, especially given the use of hearing aids and cochlear implants which increasingly enable users to accommodate and process musical acoustic signals. Difficulties regulating the loudness of playing may be the strongest single observable outcome of a musician's deafness, with different effects resulting from not being able to hear yourself (playing louder) and not being able to hear your co-performer(s) (playing quieter). In future, vibrotactile technology may be used to compensate for these effects by helping musicians with hearing impairments to regulate the loudness of their playing. There is evidence, for example, that when an auditory signal is presented simultaneously as sensory information to the skin, it is perceived to be louder by an average of 1dB (Merchel, Altinsoy, & Stamm, 2012). Similarly, Schurmann et al. (2004) found that participants matching an auditory probe and reference tone for loudness chose levels that were 12% lower when they were also touching a vibrating tube. In summary, the present study extends previous research by its authors, exploring perception thresholds and relative pitch detection in the vibrotactile domain (Hopkins, Mate-Cid, Seiffert, Fulford, & Ginsborg, 2012, 2013) and the extent to which it may be possible for people with absolute pitch to map the experience of felt vibrations to mental representations of pitch (Ginsborg, Hopkins, Fulford, Mate-Cid, & Seiffert, 2013), to suggest that the vibrations produced by auditory signals can be used to augment musicians' subjective perceptions of sound and lessen involuntary compensation effects.

5. REFERENCES

Bavelier, D., Dye, M. W. G., & Hauser, P. C. (2006). Do deaf individuals see better? Trends in Cognitive Sciences, 10*(11), 512-518. doi: 10.1016/j.tics.2006.09.006*

Bosworth, R. G., & Dobkins, K. R. (2002). The effects of spatial attention on motion processing in deaf signers, hearing signers, and hearing nonsigners. Brain and Cognition, 49*(1), 152-169. doi: 10.1006/brcg.2001.1497*

Codina, C., Pascalis, O., Mody, C., Toomey, P., Rose, J., Gummer, L., & Buckley, D. (2011). Visual advantage in deaf adults linked to retinal changes. Plos One, 6*(6). doi: 10.1371/journal.pone.0020417*

Fulford, R., & Ginsborg, J. (2013). The effects of hearing impairment on interactive performance: Two observational experiments. Paper presented at the International Symposium on Performance Science 2013, Vienna, Austria.

Fulford, R., & Ginsborg, J. (2014 in press). Can you hear me? Effects of hearing impairments on verbal and non-verbal communication during collaborative musical performance. Psychology of Music.

Fulford, R., Ginsborg, J., & Goldbart, J. (2011). Learning not to listen: the experiences of musicians with hearing impairments. Music Education Research, 13*(4), 429-446.*

Fulford, R., Ginsborg, J., & Goldbart, J. (2012, 23-28 July 2012). Functions and uses of auditory and visual feedback: Exploring the possible effects of a hearing impairment on music performance*. Paper presented at the 12th International Conference of Music Perception and Cognition and 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece.*

Ginsborg, J., Hopkins, C., Fulford, R., Mate-Cid, S., & Seiffert, G. (2013). Interactive performance: Towards the use of vibrotactile technology by musicians with hearing impairments. *Paper presented at the International Symposium on Performance Science 2013, European Association of Conservatoires, Vienna.*

Goebl, W., & Palmer, C. (2009). Synchronization of Timing and Motion Among Performing Musicians. Music Perception, 26*(5), 427-438. doi: 10.1525/mp.2009.26.5.427*

Hopkins, C., Mate-Cid, S., Seiffert, G., Fulford, R., & Ginsborg, J. (2012). Measurement of vibrotactile thresholds on the fingertip to assess vibrotactile presentation of musical notes to people with and without a hearing impairment*. Paper presented at the Proceedings of Internoise 2012, I-INCE, New York.*

Hopkins, C., Mate-Cid, S., Seiffert, G., Fulford, R., & Ginsborg, J. (2013). Inherent and learnt abilities for relative pitch in the vibrotactile domain using the fingertip. *Paper presented at the Proceedings of ICVS 20. International Congress on Sound and Vibration, Bangkok.*

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabien, R. Timmers & E. Schubert (Eds.), Expressiveness in musical performance: Empirical approaches across styles and cultures *(pp. 260-282). Oxford: Oxford University Press.*

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery and the co-ordinatory of body movements and sounds in musical ensembles. Music Perception, 28*(1), 27-46. doi: 10.1525/mp.2010.28.1.27*

Merchel, S., Altinsoy, E., & Stamm, M. (2012). Touch the sound: audio-driven tactile feedback for audio mixing applications. Journal of the Audio Engineering Society, 60*(1/2), 47-53.*

Morgan, G., Killough, C. M., & Thompson, L. A. (2011). Does visual information influence infants' movement to music? Psychology of Music. *doi: 10.1177/0305735611425897*

Proksch, J., & Bavelier, D. (2002). Changes in the Spatial Distribution of Visual Attention after Early Deafness. Journal of Cognitive Neuroscience, 14*(5), 687-701. doi: 10.1162/08989290260138591*

Rettenbach, R., Diller, G., & Sireteanu, R. (1999). Do deaf people see better? Texture segmentation and visual search compensate in adult but not in juvenile subjects. Journal of Cognitive Neuroscience, 11*(5), 560-583. doi: 10.1162/089892999563616*

Schurmann, M., Caetano, G., Jousmaki, V., & Hari, R. (2004). Hands help hearing: Facilitatory audiotactile interaction at low sound-intensity levels. Journal of the Acoustical Society of America, 115*(2), 830-832.*

Ternström, S. (2003). Choir acoustics: An overview of scientific research published to date. International Journal of Research in Choral Singing, 1*, 3-11.*

EMBODIED MEANINGS IN BEETHOVEN'S LATE PIANO SONATAS

Morton Wan[1]
[1]*The University of Hong Kong, Hong Kong, China*
Time: 09:30

Beethoven's late piano sonatas are special: abundant in meaning-laden musical gestures and epoch-making formal innovations, these sonatas seem to communicate more in symbols, ciphers, and codes than just the sound. The experience of transcendence associated with these sonatas intrigues intellectual responses that endeavor to expound what is ineffable, and such endeavors have largely been preoccupied with a methodology based primarily on third-person data. There are, however, other qualities in Beethoven's music that are as unique and intriguing as his acknowledged innovations in musical style and form. Those qualities, mostly sub-verbal and sub-intellectual, stem from a first-person, physio-psychological experience of playing the music. They give Beethoven a distinct profile both to the ears and under the hands. By way of constructing musical meanings in the process of articulating those "other" qualities, this paper attempts to introduce a new outlook on Beethoven's last three sonatas (in particular Op. 110) and consequently report some of the previously unvarnished truths about these masterworks. My method is aimed to explore the making of musical meanings through analyzing the "embodied presence" in performing the sonatas. To appropriate music as embodied experience, as I shall show, facilitates the discovery and capturing of the expression of creative desire, maximizes aesthetic immediacy, and elicits a phenomenology of music. By applying this method to Beethoven's late piano sonatas, I hope to investigate the aesthetics of mediation together with the construction of meaning in the unconventional compositional syntax of these late works. In evoking the mutually reciprocal relationship between performer and composer, the "body" operates both as the representation in a "phenomenal field" and the mediator of a "transcendental space." Through a synthesis in the twofold "body," this study also attempts to probes into the metaphysical potential in Beethoven's late sonatas, and thus serves as a phenomenological validation—from the vantage point of musical performance—of the ontological discourses about humanity that Beethoven's late sonatas seem to delineate.

PSYCHOHAPTICS AND THE AFFORDANCES OF MUSICAL INSTRUMENTS: TOWARD AN ECOLOGICAL THEORY OF PERFORMANCE

Randall Harlow[1]
[1]*University of Northern Iowa, USA*
Time: 10:00

Founded upon empirical research illuminating the psychohaptics of instrumental performance, this paper presents an ecological theory of musical performance which offers insights into the multifarious relationship between instrument design, performance practice, kinesthetics, gesture and musical meaning. By design, musical instruments present body-scaled affordances for the production of musical sound. However, conventions of performance practice, personal and cultural aesthetics, acoustics, and the social circumstances of performance are equally affordances underlying instrumental musical expression. Taken together, the affordances available to a performer constitute an ecology from which distinct performances, improvisations or compositions may emerge. Taking Eric Clarke's ecological theories of listening as a point of departure, the theory presented in this paper turns Clarke's paradigm on its head by focusing on the act of sound production. An ecological theory of instrumental performance may be grounded in the same three factors laid out by Clarke: Perception and action, adaptation, and perceptual learning. The relationship between perception and action can be seen in the psychohaptic feedback loop between kinesthetic action upon the instrument, the cognition of musical gesture, and the perception of sound; the co-evolution of musical instruments, performance techniques, and performance practice constitute an adaptive ecosystem; and perceptual learning is born out through an individual performer's development from beginner to expert. Recent music cognition and perception research evidences the epistemological reality of these three factors, pointing to an ecological model of instrumental performance. Finally, analogous to Clarke's argument for the validity of an ecological reframing of listening, this paper will examine what the reframing of instrumental performance from an ecological perspective can achieve. As examples from the history of keyboard design, performance, and literature illustrate, an ecological theory of instrumental performance offers empirical predictions, reveals previously obscured connections, and illuminates heretofore inexplicable phenomena. An ecological theory of instrumental performance unites disparate intersecting disciplines in a way which is both empirically demonstrable and intrinsically revealing.

[6C] Cognitive Modeling

MONTREAL 08:30-10:30 Thursday 7 Aug 2014

PITCH DYNAMICS ON MULTIPLE TIME SCALES IN A NEURODYNAMIC MODEL OF MELODIC PERCEPTION

Ji Chul Kim[1], *Edward W. Large*[1]
[1]*University of Connecticut, USA*
Time: 08:30

1. Background: Tonal melodies carry dynamic qualities that isolated tones do not possess. Within a tonal context, certain pitches are perceived to be stable points of repose while others are unstable and generate melodic tension that is resolved by moving to a stable pitch neighbor. Such patterns of relative tonal stability are typically found at multiple structural levels, constituting a hierarchical organization.

2. Aims: Here we attempt to model the perception of hierarchical pitch organization in tonal melodies using a neurodynamic model of pitch memory. Oscillatory neurodynamics have been shown to explain signal processing properties of the auditory system, including cochlear mechanics and brainstem frequency-following responses, as well as the perception of residue pitch, relative consonance and dissonance of intervals, and the hierarchies of tonal stability measured for Western and Hindustani music. We use a simple mathematical model of oscillatory system to build a computational model of pitch memory and show that the dynamical properties of interconnected oscillatory neural populations can explain dynamic interactions among melodic pitches on multiple time scales.

3. Method: The model consists of multiple layers of interconnected neural oscillators tuned to an equal-tempered chromatic scale. The top two layers, with fast and slow dynamics, constitute pitch memory of different time scales. Within each memory layer, coupling between oscillators gives rise to interference between pitches separated by a half or whole step as well as mode-locking resonance between pitches whose frequency ratio approximates a simple integer ratio. The activities in the memory layers are observed while the lowest layer is driven by tonal melodies made of pure tones.

4. Results: Comparison of simulated activities in the pitch memory model with the musical analysis of stimulus melodies demonstrates how hierarchical tonal organizations recognized in music theory can arise from nonlinear resonance and interference between neural oscillators. The focus of discussion will be on how mutual coupling between the fast and slow memory layers gives rise to the interactions between local and global pitch organizations.

5. Conclusions: The simulations show that multi-scale pitch dynamics in tonal melody can be explained in terms of the basic dynamical properties of oscillatory neural networks, providing a further support to the neurodynamic account of auditory perception.

TRACKING EXPRESSIVE PERFORMANCES WITH LINEAR AND NON-LINEAR TIMING MODELS

Gerald Golka[1], *Werner Goebl*[1,2]
[1]*Austrian Research Institute for Artificial Intelligence, Vienna, Austria,* [2]*Institute of Music Acoustics, University of Music and Performing Arts Vienna, Austria*
Time: 09:00

1. Background: Sensori-motor synchronization (SMS) describes the ability of humans to rhythmically coordinate movements to external stimuli. SMS can be viewed from two theoretical perspectives: According to the dynamical systems theory, SMS involves non-linear phase and period adjustments to a set of coupled internal oscillators. The information-processing approach posits linear phase and period correction of the internal timekeeper. Models derived from these theories are commonly tested in non-musical tapping experiments.

2. Aims: In music performance tempo fluctuations are used to increase expressivity which makes tracking of musical sequences more difficult for these models. In order to improve the tracking capabilities, we propose and test an extension to both models by introducing piece-specific tempo expectations. This extension is tested on a corpus of expressively performed music.

3. Methods: A set of symbolic performance data containing excerpts of Chopin's piano etude Op. 10 No. 3 performed by 22 professional pianists comprises the test corpus. Tempo expectations are modeled using local inter-onset intervals (averaged across performances). We test four

different models: 2 (model types, linear/non-linear) x 2 (information on tempo expectations, yes/no). Two methods are used: first, we optimize each model to fit individual performances so that the lowest possible timing errors (between predicted and actual note onset times) are achieved. Second, we run a cross-validation experiment to test the generalization capabilities of the models. For this purpose each model is optimized to fit a subset (60%) of the performances and the evaluation is done on the remaining performances. This procedure is repeated multiple times with random assignments to the optimization/evaluation subsets.

4. Results: The experiments are currently under way. Preliminary results suggest that adding information on typical expressive timing patterns improves model predictions.

5. Conclusions: We plan to further test these models in a musical synchronization task in which a human co-performs with a virtual duet partner driven by the extended models.

PERCEPTION OF SEGMENT BOUNDARIES IN MUSICIANS AND NON-MUSICIANS

Martin Hartmann[1], Petri Toiviainen[1], Olivier Lartillot[2]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, Finland, [2]Department of Architecture, Design and Media Technology, Aalborg University, Denmark
Time: 09:30

ABSTRACT

In the act of music listening, many people break down musical pieces into chunks such as verses and choruses. Recent work on music segmentation has shown that highly agreed segment boundaries are also considered strong and are described by using multiple cues. However, these studies could not pinpoint the effects of data collection methods and of musicianship on boundary perception. Our study investigated the differences between segmentation tasks performed by musicians in real-time and non real-time listening contexts. Further, we assessed the effect of musical training on the perception of boundaries in real-time listening. We collected perceived boundaries by 18 musicians and 18 non-musicians in 9 musical examples. Musicians also completed a non real-time segmentation task for 6 of the examples. We observed high significant correlations between participant groups and between task groups at a time-scale of 10 seconds after comparing segmentation data at different resolutions. Further, musicians located significantly more boundaries in the non real-time task than in the real-time task for 5 out of 6 examples. We found a clear effect of the task but no effects of musical training upon perceived segmentation.

1. INTRODUCTION

Music listening often prompts people to spontaneously predict and detect relevant changes that demarcate the onset and offset of verses, choruses, and other parts. This skill permits musicians and dancers to break down rehearsals into logical chunks, and helps disc jockeys and music engineers to navigate through music audio files. We can describe segmentations or boundaries in a broad sense as contrasts, discontinuities, changes and repetitions (Addessi & Caterina, 2000); in this study we will specifically refer to segmentations or boundaries as *instants of significant change* in the music (Foote, 2000), and we will focus on the high level structure instead of on phrase level segmentation.

In general, people seem to share a common sense of the time locations at which musical changes become most significant, although some people systematically tend to segment more than others (Clarke & Krumhansl, 1990; Koniari, Predazzer & Mélen, 2001; Bruderer, 2008). Experimental studies related with musical boundary perception have tried to tackle the issue of how people segment music into either an unlimited or a fixed number of parts. Other issues that have been studied include how people justify their segmentations, judge their time position, and estimate their duration. To tackle these questions, perceived boundaries were compared to perceptual interpretations (Addessi & Caterina, 2000), grouping rules (Deliège, 1987; Clarke & Krumhansl, 1990; Frankland, McAdams & Cohen, 2004), cognitive and musicological theories (Peebles, 2011), and acoustic descriptions (Bailes & Dean, 2007). Further, some studies implemented automatic segmentation systems based on musical features (Hargreaves, Klapuri & Sandler, 2012; Smith, Chuan & Chew, 2013) or on sets of rules (Lartillot & Ayari, 2009; Pearce, Müllensiefen & Wiggins, 2010), and tested the performance of the systems against the perceptual ground truth. Other related work includes a study by Burunat, Alluri, Toiviainen, Numminen and Brattico (in press) in which a perceptual segmentation task was conducted to find musical triggers of working memory, whose time locations were compared with a functional magnetic resonance imaging (fMRI) dataset obtained from participants while listening to music.

Most of the studies that investigated the effect of musicianship upon boundary perception have approached this issue from the perspective of music theory. These studies focused mainly on the differences and similarities between musician and non-musician listeners in their perception of the musical structure. Deliège (1987) assessed the segmentation of short musical stimuli by music students and non-musicians, and found similar segmentation patterns between both groups. However, music students segmented significantly more in accordance with rules of the GTTM (Generative Theory of Tonal Music, see Lerdahl and Jackendoff, 1986) than non-musicians. Bruderer (2008) collected segmentation boundaries from two subsamples of around 6 musicians and 6 non-musicians each. The stimuli used for each subsample were complete examples of Western polyphonic music in MIDI and audio versions, respectively. One of his findings was that non-musicians marked significantly more boundaries than musicians.

There is no established listening experiment method for the collection of perceived musical boundaries. However, it is usual to present the stimulus to the participant in one or more "listening only" trials, and to afterwards collect segmentation responses as participants listen again to the stimulus. This is done to ensure that the participants are familiar with the structure of the stimulus before they mark the musical boundaries. With this respect, the GTTM postulates that a complete hierarchical mental representation is only achieved after the whole example has been heard (Koniari & Tsougras, 2012). For example, Clarke & Krumhansl (1990) asked participants to segment two complete piano pieces by initially listening the complete stimulus, then marking boundaries as they listened to the music, and finally making changes or deletions of the previously marked boundaries. Bruderer (2008) asked participants to first listen the complete example to ensure familiarity with the stimulus, and then to segment as they listened once again to the stimulus. The segmentation was performed three times in a row to obtain multiple trials from the same participant. He found that the number of marked boundaries remained similar across the segmentation trials. Also Deliège (1987) utilized a familiarization phase in the listening experiments, which was followed by a segmentation task. The segmentation was offline, in the sense that participants had to mark the boundaries only after listening to the complete stimulus. In later studies a different approach was utilized; Deliège, Mélen, Stammers and Cross (1996) asked non-musician participants, in an online task, to segment a 30-second piano piece as it was listened for

the first time and found that they analyzed the music based mostly upon rhythmic and metric characteristics of the music and less upon harmonic functions. Burunat et al. (in press) asked musicians to segment large chunks of a contemporary tango piece (*Piazzolla*, which is used in the present study); each chunk was presented twice to the participants and in randomized order. Burunat has reported in a personal communication that there was a high within-subject consistency with respect to the indicated boundaries.

Our study contributes to the presented literature on some accounts; first, it introduces a *perceptual* data collection task for the comparison between musicians and non-musicians. We collected data in a real-time segmentation task that, in contrast to most of the previous studies, is not preceded by a "listening only" trial or by practice trials using the same stimuli. In this sense, it expands on the segmentation approaches by Deliège et al. (1996) and Burunat et al. (in press) since it incorporates segmentation by musicians and non-musicians, as well as diverse musical stimuli. The present study also proposes a *cognitive* data collection task; this task resembles the approach by Clarke and Krumhansl (1990) but replaces the tools utilized for fine-tuning the position of the marked boundaries. Instead of a score-based annotation, we obtained precise annotations using audio editing software, similarly to the approach by Wiering, de Nooijer, Volk and Tabachneck-Schijf (2009) for segmentation of rendered MIDI melodies. In short, our contribution to the state of the art with respect to the data collection task could help to understand what are the differences between an immediate hierarchical representation of the music and a perhaps more meaningful description obtained after a complete listen of the stimulus and the possibility to reconsider the location and marking of the boundaries. In addition, we chose a subsample that is almost half the size of the original sample because we think that the dividing line between musicians and non-musicians is fuzzy, and therefore we wanted to look at the extremes of the distribution. For example, we included in the subsample only those musicians who self-reported themselves as semiprofessional or professional musicians. Bruderer (2008) also utilized a subsample of the participants, but his findings are based only on relatively small groups of non-musicians and non-professional musicians. Other studies did not provide enough information about the musicianship levels of the subjects. Moving to a more technical viewpoint, our data analysis is based on Kernel Density Estimation (KDE, Silverman, 1986) matrices for data representation, extends the number and range of considered time-scale resolutions with respect to previous work and presents an alternative approach for the calculation of an optimal time-scale.

We believe that the study of music segmentation can deepen our knowledge of temporal processing of perceptual streams, which are found in music, speech and movement, and that the study of the effects of musicianship upon musical boundary perception can help us gain a better understanding of the possible transfer effects of music learning. In addition, systematic studies on the perception of instants of significant musical change can encourage developments in automatic segmentation tools to facilitate music editing and playback for, among others, everyday life tasks such as adding music to family videos.

People seem to intuitively understand music as a conglomerate with relatively precise boundaries. Further, many subjects may indicate musical boundaries at similar locations (Bruderer, Mckinney & Kohlrausch, 2009). There even seems to be a relationship between the number of subjects that assigned a boundary to a given time location and its rated salience (Bruderer, 2008). Musical boundaries cannot be arbitrary, and might instead emerge as a complex interplay between distinct musical structures that iterate or variate throughout a piece and our psychological mechanisms of perception and cognition (Deliege, 2007). There should be manifest and contrasting events in the music that prompt people to perceive beginnings and ends of musical segments. However, such events could differ in their number and time locations based upon multiple factors. For example, there can be differences in the musical training level of the listeners and in the data collection method that is used to gather segment boundaries. We primarily aimed to reach further insights regarding these two aforementioned factors. Our aims could be condensed into the following research questions:

1. What is the effect of musicianship on the perception of musical segment boundaries?

2. What are the differences between a spontaneous first impression of the musical structure as it unfolds over time and a deeper, more knowledge-driven impression?

We predicted that musicians would segment differently from non-musicians, perhaps because of differences in the perception of e.g. harmonic changes in the music. We also estimated that participants would segment differently in each task, since they would indicate more surface changes, such as rhythmic changes, in the perceptual (or real-time) task than in the cognitive task. We estimated the cognitive (or non real-time) task data to parse the musical structure based upon deeper changes.

2. EXPERIMENT I – PERCEPTUAL TASK

We conducted two listening experiments on perceived segmentation, where the first was a prerequisite for the second. In the first experiment, participants were asked to indicate significant musical boundaries at the same time as they listened to unfamiliar musical examples. The aim was to capture a fresh, "live" description or first impression of the music as it unfolded over time.

We collected real-time segmentation responses from participants using computers with a Max/MSP Patch. The stimuli comprised 9 musical pieces of a variety of Western musical styles. We included an Appendix to this paper with a glossary of the abbreviations used for the stimuli and with information about their duration. The music was played back to the participants using headphones at a comfortable volume level. We originally collected segmentation data from 74 participants, and later chose a final sample that comprised 18 non-musicians (11 males, 7 females) and 18 musicians (10 females, 8 males). The mean age of the participants was 27.45 years. They were all students or graduates from different faculties of the University of Jyväskylä and of the JAMK University of Applied Sciences. The musicians had an average of 14.39 years of musical training. All the non-musicians reported having had no musical training, whereas all of the selected musicians considered themselves either as semiprofessional or professional musicians at the time of the data collection.

The experiment took place with a computer in a sound-attenuated room. The participants were instructed to mark instants of significant change as they listened to the music by pressing the space bar of the computer keyboard. After completing a trial, they listened and marked each of the musical stimuli, which were presented in a randomized order. Participants were instructed to give their "first impression" because they would not have a chance to listen to the whole example before they started marking. The interface included a play bar that offered basic visual-spatial cues regarding the beginning, current time position and end of the examples. On average, it took 47 minutes to complete the whole experiment for the 18 non-musicians chosen for the study and 50 minutes for the 18 musicians. The participants filled a questionnaire at the end of the experiment, which included demographic and musicianship questions.

3. EXPERIMENT II – COGNITIVE TASK

The second experiment was conducted in order to obtain, from the musician participants, a segmentation that would be more comprehensive and precise than the first one. We prepared an interface in Sonic Visualiser to collect segmentation boundaries and perceived strength from musical examples. We intended to keep the duration of this experiment at around one hour, so we chose 6 examples from Experiment I that lasted around 2 minutes each for Experiment II. Headphones were used to playback the music at a comfortable listening level.

The final sample consisted of 18 musician participants (10 females, 8 males); they were selected among 36 subjects who completed the second experiment. All the participants had previously taken part of the first experiment. We did not recruit non-musicians for Experiment II because only a few of them reported experience in audio editing.

This experiment took also place in a sound-attenuated room with a computer. Exceptionally, five subjects participated at the same time in a sound-attenuated classroom with computers. Two of them were chosen for the final sample of 18 participants. Compared to Experiment I, the second experiment required the training to be completed while the experimenter was in the room. The experimenter read the instructions together with the participant and presented the interface. He asked the participant to perform the task upon two short trial stimuli by following the experiment instructions. Once the trial concluded, the experimenter left the room and the participant could start with the task. Participants were asked to: 1) Listen to the complete musical example; 2) Listen again to the complete example, and at the same time mark instants of significant change by pressing the Enter key; 3) Freely playback the musical example from different time points and correct marked positions to make them more precise, or remove them if these were added by mistake; they were also asked not to add any new markings at this stage; 4) Mark the strength of the significant change for each instant with a value ranging from 1 to 10; 5) Move to the next musical example and start over from the first step. The interface showed participants the waveform of the musical examples, over which they would play back and segment the stimuli, correct the boundaries and mark their strength. The participants were asked to focus on the music and not on the visual content. It took the 18 chosen participants one hour in average to complete the second experiment.

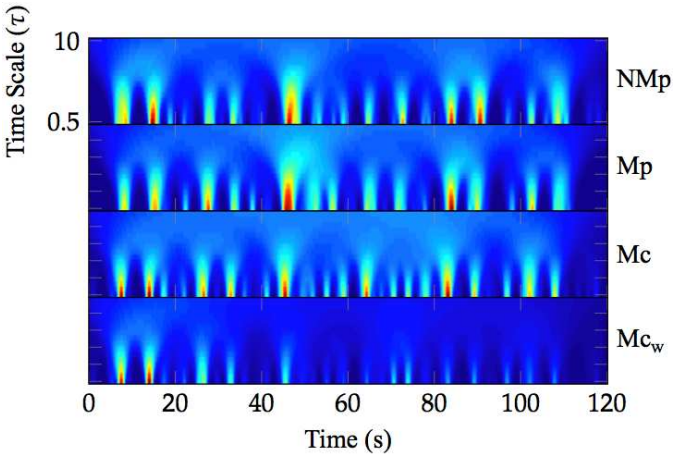


Figure 1: The segmentation sets of data marked by participants are visualized as multi-resolution KDE matrices for the musical example *Ragtime*. The density function over time is represented within each of the four matrices and for each time-scale of the KDE considered. Warm colors denote high values while cool colors denote low values.

4. RESULTS

We organized the segmentation responses into three main groups based on the level of musicianship of the participants and the corresponding segmentation task. We allocated 162 segmentations to the group of musicians in the perceptual task, 162 segmentations to the group of non-musicians in the perceptual task, and 108 segmentations to the group of musicians in the cognitive task. We abbreviated these groups as *NMp* for non-musicians in the perceptual task, *Mp* for musicians in the perceptual task, and *Mc* for musicians in the cognitive task. We chose a method to visualize the segmentations that would summarize the segmentation data in a precise and hierarchical way. We added together the responses of each group and task using KDE's with different smoothing bandwidths. For each song, we obtained 16 KDE curves that were organized into matrices following a multi-resolution approach that has been previously utilized upon musical descriptors (Martorell Dominguez, 2013) and novelty curves (Kaiser & Peeters, 2013). Since we had collected ratings from participants on perceived boundary strength in the second experiment, we included a fourth group of segmentation responses. This group corresponded to the responses by musicians in the cognitive task with added boundary strength weights, and it is abbreviated as *Mcw*. The boundary indications in the fourth group were at the same time positions as in Mc; this would allow us to estimate the effect of adding perceived boundary strength to the cognitive segmentations. We obtained 30 KDE matrices, since the perceptual task data (NMp and Mp) was based on 9 stimuli and the cognitive task data (Mc and Mcw) was based on 6 stimuli. Figure 1 shows the 4 KDE matrices that were obtained for the example Ragtime. We constructed the matrices after obtaining KDE curves at different KDE time-scales (τ). We considered 16 time-scales logarithmically ranging from .5 seconds to 10 seconds, based upon the time-span of the working memory. We did not choose linearly spaced KDE time-scales since we followed the assumption, in agreement with Weber's law, that time is perceived on a logarithmic-like scale.

5. ANALYSIS

We found a higher mean number of indicated boundaries across participants in Mc (11.33) than in Mp (5.8) for these musical examples. We computed paired samples, two-tailed *t*-tests to estimate the significance level of the difference between Mp and Mc with respect to the number of indicated boundaries by musicians. The differences reached significance for 5 out of 6 examples: at $p < .01$ for the example *Couperin* (paired $t(17)$

= 3.3, p < .01), and at $p < .05$ for the examples *Genesis* (paired $t(17) = 2.3$, $p < .05$), *Smetana* (paired $t(17) = 2.1$, $p < .05$), *Ragtime* (paired $t(17) = 2.8$, $p < .05$) and *Ravel* (paired $t(17) = 2.2$, $p < .05$).

To look further at the similarity between the segmentation groups, we calculated the correlation between pairs of KDE matrices for each stimulus. As shown in Figure 2, we found strong correlations for all stimuli between Mp and NMp, and these correlations were significant at $p < .001$ based on a Montecarlo simulation with 10000 iterations. Similarly, we found strong significant correlations ($p < .001$) between KDEs corresponding to the Mc and Mcw for all musical examples. In addition, we found moderately strong correlations between the KDEs of Mp and Mc, but these only reached statistical significance ($p < .001$) for the *Dvorak* stimulus. We also found moderately strong significant correlations ($p < .001$) between Mp and the Mcw for the same stimulus.

We also compared, for each musical example, the KDE curves that were obtained with different bandwidths in order to find an optimal time-scale for segmentation responses. The KDE's were compared between pairs of sets of data using the identical smoothing parameters. The correlation between musicians and non-musicians for the perceptual task (Mp-NMp) was at least high for all the considered time-scales. In contrast, the comparisons between Mp and Mcw mostly exhibited moderate correlations. The comparisons between Mp and Mc indicated a general tendency to increase from low correlations at the lowest time-scales to high correlations at time-scales higher than 2 seconds. We also obtained very high correlations at all time-scales with a tendency to gradually decrease at higher time-scales for the comparison between Mc and Mcw. This decreasing tendency was much more important for *Couperin* and *Ragtime*, which correspond to solo piano performances.

We calculated a mean curve across all stimuli and considered the time-scale with the maximum correlation value of the mean curve as an optimal KDE time-scale for comparison between groups. We found a strong ($\rho = .89$, $t < .001$) overall maximal correlation coefficient for the comparison of Mp and NMp at an optimal time-scale of 10 seconds.

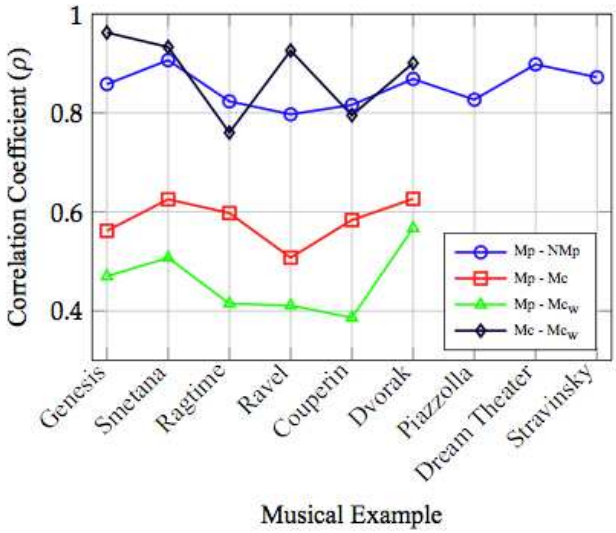


Figure 2: Profiles showing the correlation between multi-resolution KDE matrices.

6. DISCUSSION

We obtained a notably higher mean number of perceived boundaries by musicians in the cognitive task than in the perceptual task. The participants marked more boundaries in the cognitive task for all six musical examples, and for five of them, the trend reached statistical significance, suggesting an effect of the data collection task upon the number of indicated boundaries. We could argue that the differences between task groups are related with the progressive familiarization with the stimuli, since the participants had already listened to the stimuli once in the perceptual task and they were asked to listen to the complete stimuli once again before the segmentation task of the cognitive experiment. This would give support to the idea that a thorough hierarchical mental representation of a musical piece can only be reached once it has been listened in its entirety (Koniari & Tsougras, 2012). Participants might have also noticed more boundaries in the perceptual task than those that they actually marked. Alternatively, some boundaries could have been perceived but left unmarked in order to avoid markings located after the occurrence of the boundary. It could have also been the case that the cognitive task gave margin to indicate boundaries after these were perceived since these could have been later repositioned to previous time instants.

In addition, we found strong and significant correlations between the cognitive task with and without the addition of boundary strength. We found however that *Couperin* and *Ragtime*, two solo piano performances, yielded lower correlations than the other examples. Musicians rated the boundaries of these examples with relatively low strength, which increased the difference between the cognitive task sets. These two musical examples are characterized by relatively less timbral contrasts than other musical examples, which probably prompted lower boundary strength ratings. Other musical examples have strong changes in harmony, dynamics and rhythm that are accompanied with percussion instruments or with changes in instrumentation. Even *Ravel*, which is not multi-instrumental but a piano piece instead, sounds very different from *Couperin* and *Ragtime* with its contrasting melodic passages and sudden changes in register, harmony, dynamics and rhythmic patterns. We also noticed that, for both *Ragtime* and *Couperin*, participants indicated relatively important strength in the beginning as the main themes were introduced, but the boundaries that they indicated for variations of these themes were rated with lower strength. We assume that this contrast between parts in the overall strength of the examples led to a lower correlation than if the strength markings had been more homogeneous.

We found a strong significant correlation between KDE matrices corresponding to segmentation by musicians and non-musicians for the perceptual task, contradicting previous findings by Bruderer (2008) who found effects of musical training upon the indication of musical boundaries. Another finding was that the optimal time-scale for comparison of segmentations by musicians and non-musicians corresponds to the maximal time-scale considered (10 s). This raises the question of whether the range of time-scales considered could be extended in order to find out if the optimal time-scale of the examples exceeds 10 seconds. We doubt, however, that the perception of segment boundaries would be appropriately represented if we computed KDEs at smoothing bandwidths that exceeded the temporal span of working memory.

7. CONCLUSIONS

We compared perceived segmentation between musicians and non-musicians in a real-time segmentation task. We did not find evidence of an effect of musical training upon music segmentation using the proposed approach, since we found similar segmentation profiles between both groups and a similar number of marked boundaries. We also found an overall maximal correlation between these two groups at a time-scale of 10 seconds. In addition, we found that musicians marked significantly more boundaries for the same stimuli in the cognitive segmentation task than in the perceptual segmentation task, which might be due to an increased familiarity with the stimuli or to other differences between the data collection tasks. Our results showed that relatively large time-scales, corresponding to a high-level hierarchy of the musical structure, are optimal for comparison of segmentation responses between musicians and non-musicians, and may be appropriate parameters for representation of perceived musical change. To gain more understanding on optimal KDE time-scales, further work could focus on alternatives to fixed smoothing bandwidths such as variable KDE estimation methods. Future work could focus on which specific boundaries are indicated in one task but not in the other one, and explore preliminary and final time positions of boundaries in the cognitive task. We will attempt to gain more insights regarding the issue of segmentation in future work by assessing the relationship between perceived segmentation and quantitative musical descriptions extracted from the stimuli.

8. REFERENCES

Addessi, A. R. and Caterina, R. (2000). Perceptual musical analysis: Segmentation and perception of tension. Musicæ Scientiæ, *4(1):31–54.*

Bailes, F. and Dean, R. T. (2007). Facilitation and coherence between the dynamic and retrospective perception of segmentation in computer-generated music. Empirical Musicology Review, *2(3).*

Bruderer, M. J. (2008). Perception and Modeling of Segment Boundaries in Popular Music. *PhD thesis, JF Schouten School for User-System Interaction Research, Technische Universiteit Eindhoven, Netherlands.*

Bruderer, M. J., Mckinney, M. F., and Kohlrausch, A. (2009). The perception of structural boundaries in melody lines of western popular music. Musicæ Scientiæ, *13(2):273–313.*

Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (in press). Dynamics of brain activity underlying working memory for music in a naturalistic condition. Cortex. *DOI:10.1016/j.cortex.2014.04.012*

Clarke, E. and Krumhansl, C. (1990). Perceiving musical time. Music Perception, *pages 213–251.*

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. Music Perception, *pages 325–359.*

Deliège, I., Mélen, M., Stammers, D., and Cross, I. (1996). Musical schemata in real-time listening to a piece of music. Music Perception, *pages 117–159.*

Frankland, B. W., McAdams, S., and Cohen, A. J. (2004). Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's a generative theory of tonal music. Music Perception, *21(4): pp. 499–543.*

Hargreaves, S., Klapuri, A., and Sandler, M. (2012). Structural segmentation of multitrack audio. IEEE Transactions on Audio, Speech, and Language Processing, *20(10).*

Kaiser, F. and Peeters, G. (2013). Multiple hypotheses at multiple scales for audio novelty computation within music. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, *Vancouver.*

Koniari, D., Predazzer, S., and M´elen, M. (2001). Categorization and schematization processes used in music perception by 10-to 11-year-old children. Music Perception, *18(3):297–324.*

Koniari, D. and Tsougras, C. (2012). The cognition of grouping structure in real-time listening of music. a GTTM-based empirical research on 6 and 8-year-old children. In Proceedings of the 12th International Conference on Music Perception and Cognition, *Thessaloniki, Greece.*

9. APPENDIX – LIST OF STIMULI

Genesis – Banks, T., Collins, P. & Rutherford, M. (1986). The Brazilian. [Recorded by Genesis]. On *Invisible Touch* [CD]. Virgin Records. (1986) Excerpt: 01:10.200-02:58.143

Smetana – Smetana, B. (1875). Aus Böhmens Hain und Flur. [Recorded by Gewandhausorchester Leipzig - Václav Neumann]. On *Smetana: Mein Vaterland* [CD]. BC - Eterna Collection. (2002) Excerpt: 04:06.137-06:02.419

Ragtime – Morton, F. (1915). Original Jelly Roll Blues. On *The Piano Rolls* [CD]. Nonesuch Records. (1997) Excerpt: 0-02:00.104

Ravel – Ravel, M. (1901). Jeux d'Eau. [Recorded by Martha Argerich]. On *Martha Argerich, The Collection, Vol. 1: The Solo Recordings* [CD]. Deutsche Grammophon. (2008) Excerpt: 03:27.449-05:21.884

Couperin – Couperin, F. (1717). Douzième Ordre / VIII. L'Atalante. [Recorded by Claudio Colombo]. On *François Couperin : Les 27 Ordres pour piano, vol. 3 (Ordres 10-17)* [CD]. Claudio Colombo. (2011) Excerpt: 0-02:00

Dvorak – Dvořák, A. (1878). Slavonic Dances, Op. 46 / Slavonic Dance No. 4 in F Major. [Recorded by Philharmonia Orchestra - Sir Andrew Davis]. On *Andrew Davis Conducts Dvořák* [CD]. Sony Music. (2012) Excerpt: 00:57.964-03:23.145

The following examples were used for Experiment I only. We trimmed these ~8 minute examples into sections of ~2 minutes each for a more even length distribution across the pool of stimuli and to avoid fatigue to the participants. The sections were overlapped by 3 seconds, which

corresponds to the duration of the echoic memory store. We later concatenated the segmentation data in order to obtain a set of indicated boundaries for the complete stimulus. We corrected the overlapping segmentation data by discarding the first 3 seconds of each non-initial chunk.

Piazzolla – Piazzolla, A. (1959). Adiós Nonino. [Recorded by Astor Piazzolla y su Sexteto]. On *The Lausanne Concert* [CD]. BMG Music. (1993) Excerpt: 0-08:07.968. Trimmed sections: 0-02:00, 01:57-03.57, 03:54-05:54, 05:51-08:07.968

Dream Theater – Petrucci, J., Myung, J., Rudess, J. & Portnoy, M. (2003). Stream of Consciousness (instrumental). [Recorded by Dream Theater]. On *Train of Thought* [CD]. Elektra Records. (2003) Excerpt: 0-07:50.979. Trimmed sections: 0-02:00, 01:57-03.57, 03:54-05:54, 05:51-07:50.979

Stravinsky – Stravinsky, I. (1947). The Rite of Spring (revised version for Orchestra) Part I: The Adoration of The Earth (Introduction, The Augurs of Spring: Dances of the Young Girls, Ritual of Abduction). [Recorded by Orchestra of the Kirov Opera, St. Petersburg - Valery Gergiev]. On *Stravinsky: The Rite of Spring / Scriabin: The Poem of Ecstasy* [CD]. Philips. (2001) Excerpts: 00:05-03:23, 0-03:12, 0-01:16 - total duration: 07:47.243. Trimmed sections: 00:05-02:05, 02:02-04:02, 03:59-05:59, 05:56-07:52.243

Lartillot, O. and Ayari, M. (2009). Segmentation of Tunisian modal improvisation: Comparing listeners' responses with computational predictions. Journal of New Music Research, *38(2):117–127.*

Lerdahl, F. and Jackendoff, R. (1986). A generative theory of tonal music. *The MIT Press, Cambridge, M.A.*

Martorell Dominguez, A. (2013). Modelling Tonal Context Dynamics by Temporal Multi-Scale Analysis. *PhD thesis, Universitat Pompeu Fabra, Barcelona.*

Pearce, M. T., Müllensiefen, D., and Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. Perception, *39(10):1367–1391.*

Peebles, C. (2011). The Role Of Segmentation And Expectation In The Perception Of Closure. *PhD thesis, Florida State University, Florida.*

Silverman, B. W. (1986). Density estimation for statistics and data analysis. *London: Chapman and Hall.*

Smith, J. B. L., Chuan, C., and Chew, E. (2013). Audio properties of perceived boundaries in music. IEEE Transactions on Multimedia.

Wiering, F., de Nooijer, J., Volk, A., and Tabachneck-Schijf, H. (2009). Cognition-based segmentation for music information retrieval systems. Journal of New Music Research, *38(2):139–154.*

APPENDIX – LIST OF STIMULI

Genesis – *Banks, T., Collins, P. & Rutherford, M. (1986). The Brazilian. [Recorded by Genesis]. On* Invisible Touch *[CD]. Virgin Records. (1986) Excerpt: 01:10.200-02:58.143*

Smetana – *Smetana, B. (1875). Aus Böhmens Hain und Flur. [Recorded by Gewandhausorchester Leipzig - Václav Neumann]. On* Smetana: Mein Vaterland *[CD]. BC - Eterna Collection. (2002) Excerpt: 04:06.137-06:02.419*

Ragtime – *Morton, F. (1915). Original Jelly Roll Blues. On* The Piano Rolls *[CD]. Nonesuch Records. (1997) Excerpt: 0-02:00.104*

Ravel – *Ravel, M. (1901). Jeux d'Eau. [Recorded by Martha Argerich]. On* Martha Argerich, The Collection, Vol. 1: The Solo Recordings *[CD]. Deutsche Grammophon. (2008) Excerpt: 03:27.449-05:21.884*

Couperin – *Couperin, F. (1717). Douzième Ordre / VIII. L'Atalante. [Recorded by Claudio Colombo]. On* François Couperin : Les 27 Ordres pour piano, vol. 3 (Ordres 10-17) *[CD]. Claudio Colombo. (2011) Excerpt: 0-02:00*

Dvorak – *Dvořák, A. (1878). Slavonic Dances, Op. 46 / Slavonic Dance No. 4 in F Major. [Recorded by Philharmonia Orchestra - Sir Andrew Davis]. On* Andrew Davis Conducts Dvořák *[CD]. Sony Music. (2012) Excerpt: 00:57.964-03:23.145*

The following examples were used for Experiment I only. We trimmed these ~8 minute examples into sections of ~2 minutes each for a more even length distribution across the pool of stimuli and to avoid fatigue to the participants. The sections were overlapped by 3 seconds, which corresponds to the duration of the echoic memory store. We later concatenated the segmentation data in order to obtain a set of indicated boundaries for the complete stimulus. We corrected the overlapping segmentation data by discarding the first 3 seconds of each non-initial chunk.

Piazzolla – *Piazzolla, A. (1959). Adiós Nonino. [Recorded by Astor Piazzolla y su Sexteto]. On* The Lausanne Concert *[CD]. BMG Music. (1993) Excerpt: 0-08:07.968. Trimmed sections: 0-02:00, 01:57-03.57, 03:54-05:54, 05:51-08:07.968*

Dream Theater – *Petrucci, J., Myung, J., Rudess, J. & Portnoy, M. (2003). Stream of Consciousness (instrumental). [Recorded by Dream Theater]. On* Train of Thought *[CD]. Elektra Records. (2003) Excerpt: 0-07:50.979. Trimmed sections: 0-02:00, 01:57-03.57, 03:54-05:54, 05:51-07:50.979*

Stravinsky – *Stravinsky, I. (1947). The Rite of Spring (revised version for Orchestra) Part I: The Adoration of The Earth (Introduction, The Augurs of Spring: Dances of the Young Girls, Ritual of Abduction). [Recorded by Orchestra of the Kirov Opera, St. Petersburg - Valery Gergiev]. On* Stravinsky: The Rite of Spring / Scriabin: The Poem of Ecstasy *[CD]. Philips. (2001) Excerpts: 00:05-03:23, 0-03:12, 0-01:16 - total duration: 07:47.243. Trimmed sections: 00:05-02:05, 02:02-04:02, 03:59-05:59, 05:56-07:52.243*

COMPUTATIONAL MODEL-BASED ANALYSIS OF CONTEXT EFFECTS ON CHORD PROCESSING

Satoshi Morimoto[1], Gerard B. Remijn[2], Yoshitaka Nakajima[2]
[1]Graduate School of Design, Kyushu University, Japan, [2]Faculty of Design, Kyushu University, Japan

Time: 10:00

Musical expectancy is one of the important factors in music perception. The patterns of expectancy in chord progression have been arranged in Western music theory, as practical rules to compose tonal music. Previous studies have shown that chord priming is consistent with the music theory and is based on automatic activation processes of tonal knowledge. However, the computational process to build up musical expectancy from the preceding context has not been clarified yet. Since the music theory does not cover the entire perceptual pattern of chord priming, including atonal combinations, analyses based on the music theory are not enough to understand the context effects. Meanwhile, studies in the field of music informatics have shown that stochastic models (e.g. a Bayesian model) performed well for automatic key finding problems. These findings indicate the possibility that humans also assign a similar stochastic framework to musical contexts. We examined the context effects on listeners' chord processing from a computational perspective, using stochastic modeling, and investigated the relationship between the individual

perceptual processes and Western music theory. We conducted a behavioral experiment, in which participants listened to chord sequences and evaluated how well the last chord of each sequence belonged to the preceding sequence perceptually. The chords in each sequence were 12 major triads of 0.5 s. The sequences consisted of 2, 3 or 4 chords. We built stochastic models which simulated the computational process governing the relationship between the chord sequences and behavioral responses. We compared validity of different models for each participant by calculating likelihoods. The participants' judgments were best approximated by a model which holds multiple patterns of chord expectancy, hierarchically organized for each chord as a perceptual center; the likelihoods of the patterns are assumed to be updated sequentially by observations. The estimated patterns showed a common tendency, but were not always matched among participants. Our results suggested that, when listening to chord sequences, internally-constructed tonal assumptions dominate the musical expectancy for the subsequent chord. The patterns of context effects were different between individuals.

[6D] Absolute Pitch and Language

LIÈGE 08:30-10:30 Thursday 7 Aug 2014

SPLIT BETWEEN MUSICAL PITCH AND SPEECH PITCH PROCESSING, EVIDENCE FROM EARLY INFANCY

Ao Chen[1], *Rene Kager*[1]
[1]*Utrecht University, Netherlands*

Time: 08:30

It has been a long lasting debate whether language and music share the same neuro-cognitive resources or whether they are independent domains in a modular way (e.g. Peretz, Champod & Hyde, 2003; Fedorenko, Behr, & Kanwisher, 2011; Wong & Perrachione, 2007, among others). To shed light on this question, we tested 4-month-old infants, who have not built up specific knowledge of the native language yet, on their discrimination of pitch patterns realized on speech syllables and realized as musical melodies. The visual fixation paradigm was adopted. In this paradigm, the infants were first habituated with one sound until their attention dropped below a preset criterion. Once the habituation criterion was met, the test phase started. In the test phase, they were presented with one "old trial", in which they heard the same sound as they had heard in the habituation phase, followed by one "novel trial", in which a new sound was presented. If the infant were able to discriminate between the two sounds, then there should be a significant increase in listening time upon hearing the "novel trial", due to the recovery of interests. In the speech task, the infants were to discriminate between a rising tone and a dipping tone realized on the syllable /ma/, with a duration of 450ms. The offset of the two pitches had a difference of 5 semitones. In the music task, the infants were to discriminate two 450ms long three-note melodies DEF and DCF on middle C, which share the pitch contours with the speech stimuli. No significant discrimination effect was found in either task. However, once the two melodies were extended to 830ms, which is more typically musical according to adults' report, the infants succeeded in discriminating them. Moreover, with a duration of 830ms, the infants succeeded in discriminating DEF and #DF#F, which only differ in one semitone in absolute pitch height. Our results demonstrate that as early as 4 months, when music and speech are presented prototypically, infants are far more acute at perceiving musical pitch, which suggests early split in processing between the two domains.

A DEVELOPMENTAL STUDY OF LATENT ABSOLUTE PITCH IN UK CHILDREN

Kelly Jakubowski[1], *Daniel Müllensiefen*[1], *Lauren Stewart*[1]
[1]*Goldsmiths, University of London, United Kingdom*

Time: 09:00

1. Background: The ability to recall the absolute pitch level of familiar music (latent absolute pitch (AP)) has been found to be widespread in adults, in contrast to the rare ability of labelling or producing single pitches without a reference tone (overt absolute pitch). Previous research suggests evidence for absolute processing of pitch even in infancy, however conflicting theories and evidence exist as to whether latent AP is heightened during childhood or develops during a critical period similarly to overt AP.

2. Aims: The study aimed to investigate the developmental profile of latent absolute pitch in UK children. The second aim was to clarify previous mixed findings in this area by exploring individual differences in latent AP ability related to age, regularity of exposure to the musical stimuli, and musical training.

3. Method: Two experiments tested the latent AP ability of 288 children ages 4-12 years. On each trial, participants heard two versions of a musical excerpt from a familiar TV program differing only in their musical key. Participants were asked to choose which version sounded most like what they heard on TV at home. The two experiments differed only in the amount of pitch shifting employed (either 1 or 2 semitones from the original key).

4. Results: The sample of UK children as a whole performed significantly above chance at recognizing the absolute pitch level of familiar melodies in both experiments. The probability of success was 0.541 for the one semitone experiment ($p < .05$; 95% CI [0.51, 0.57]) and 0.580 for the two semitone experiment ($p < .001$; 95% CI [0.53, 0.63]). Mixed effects and conditional tree models were employed to model additional factors which might explain individual differences in task performance. However, individual differences between participants could not be explained by age, gender, musical training, or regularity of exposure to the stimuli.

5. Conclusions: The present findings suggest latent AP as a stable ability existing throughout childhood into adulthood. These findings have some parallels to the literature on latent AP in adulthood, in which latent AP also appears to be independent of both musical training and familiarity.

ABSOLUTE PITCH AMONG STUDENTS AT THE SHANGHAI CONSERVATORY OF MUSIC: A LARGE-SCALE DIRECT-TEST STUDY

Xiaonuo Li[1,2]
[1]*Diana Deutsch, USA,* [2]*Jing Shen, USA*

Time: 09:30

This paper reports a large-scale direct-test study of absolute pitch (AP) in students at the Shanghai Conservatory of Music. Overall performance levels were very high, and there was a substantial advantage to early onset of musical training. Students who had begun training at age ≤5 scored 83% correct not allowing for semitone errors, and 90% correct allowing for semitone errors. Performance levels were higher for white key pitches than for black key pitches. This effect was greater for orchestral performers than for pianists, indicating that it cannot be attributed to early training on the piano. Rather, accuracy in identifying notes of different names (C, C#, D, etc.) correlated with their frequency of occurrence in a large sample of music taken from the Western tonal repertoire. There was also an effect of pitch range, so that performance on tones in the 2-octave range beginning on Middle C was higher than on tones in the octave below Middle C. In addition, semitone errors tended to be on the sharp side. Further, the evidence ran counter to the hypothesis, previously advanced by others, that the note A plays a special role in pitch identification judgments.

THE EFFECT OF MUSICAL PITCH ON THE PERCEPTION OF LEXICAL PITCH IN SPEAKERS OF MANDARIN.

Joey L. Weidema[1], M. Paula Roncaglia-Denissen[1], Henkjan Honing[1]
[1]Music Cognition Group, Institute for Logic, Language and Computation (ILLC), Amsterdam Brain and Cognition (ABC), University of Amsterdam, Netherlands
Time: 10:00

1. Background: Recent studies suggest that pitch processing in language and music relies on domain-general cognitive mechanisms, modulated by experience-dependent processing (Perrachione et al., 2013; Bidelman, Hutka & Moreno, 2013; Chandrasekeran, Krishnan & Gandour, 2009; Pfordresher & Brown, 2009; but see Peretz & Coltheart, 2003; Peretz, 2012 for a modular approach). Tonal languages such as Mandarin provide the opportunity to investigate both melodic and linguistic pitch processing as perception of pitch contour of a sentence shows parallels with properties of a musical melody. Studies on pitch processing in music and language have focused on the lexical level (e.g. Marie et al., 2011) while the sentence level remains unexplored. Fully understanding pitch processing requires one to look beyond the lexical level where different phonological rules interact. This may lead to a better understanding of cognitive mechanisms governing pitch in language and in how far these are shared with those of music.

2. Aim: The current study aims to investigate the effect of pitch in a musical context on the perception of lexical tone. By exploring whether pitch in music facilitates the processing of pitch in language, behavioural data will provide evidence for or against shared cognitive mechanisms of pitch processing in these domains.

3. Method: Mandarin speakers are presented with single sentences containing their original pitch contour, or with manipulated sentences in which the last word (critical item) has a flattened lexical tone that is agrammatical. Sentences are accompanied by single melodic phrases created by transposing the pitch contour of the sentences onto a tempered scale. The relative distance between pitches is preserved, creating either a deviant or a congruent pitch with the original, non-manipulated last lexical tone. Natural-contoured sentences are used as fillers, presented with and without melody. Participants perform a content-verification task where accuracy rates and reaction times are measured.

4. Results and Conclusions: If lexical tone identification is facilitated by pitch in music then faster reaction times and higher accuracy rates are expected when the pitch contour of a melody and a sentence are congruent. This will be evidence in support of shared resources governing pitch processing in music and language. However, if no facilitation is found during the simultaneous processing of lexical and melodic tone, then the evidence will be in support of a distinct, rather than domain-general module underlying pitch processing in language and music.

Keynote Address 3 & Young Researcher Award 1

SEOUL 11:00 – 12:30 Thursday 7 Aug 2014

MUSIC, EMPATHY AND CULTURAL UNDERSTANDING

Eric Clarke
University of Oxford
Time: 11:00

Music is a source of intense experiences of both the most individual and massively communal kind, and it increasingly brings together – or exploits – an exceptional range of cultures and histories. It is not uncommon to find claims being made for music's capacity to overcome (even transcend) cultural difference, and to break down barriers of ethnicity, age, social class, disability, and physical and psychological ill health. In recent years, a broad notion of empathy has gained considerable attention and currency across musicology, psychology of music, sociology of music and ethnomusicology as a way to conceptualize a whole range of affiliative and identity-constructing capacities in relation to music. But what is brought together or understood by the term 'empathy', and is it a useful and coherent way to think about music in relation to its individual and social effects? This paper addresses the disparate nature of the evidence for the claims about music's transformative power, individually and socially, across a wide disciplinary range of theories and findings. From research on music and mirror neurons, to the ethnomusicology of affect, the history of musical subjectivity, and sociological studies of music and collective action, the case has been made for different perspectives on music's capacity to afford compassionate and empathetic insight and affiliation, and its consequent power to change social behaviour. These diverse research strands all point to the crucial role that musicking plays in people's lives, to its transformational capacity, and to the insights that it can afford. There is no single window onto 'what it is like to be human', but musicking seems to offer as rich, diverse, and globally distributed a perspective as any – and one that engages people in a vast array of experiences located along dimensions of public and private, solitary and social, frenzied and reflective, technological and bodily, conceptual and immediate, calculated and improvised, instantaneous and timeless. While a wide range of studies have suggested that empathetic interaction with other human beings is facilitated by musical engagement, the direct empirical evidence for this important possibility is scattered and disciplinarily disconnected. This paper aims to bring together a cross-section of some this evidence, to examine some of the claims in a critical light, and to propose a framework within which to understand these disparate elements so as to reveal points of interdisciplinary convergence and divergence.

Young Researcher Award 1

THE EFFECTS OF ANOMALY ON MUSIC READING: EVIDENCE FROM EYE MOVEMENTS

Lauren Hadley[1], *Patrick Sturt*[1], *Martin Pickering*[1]
[1]*University of Edinburgh, United Kingdom*
Time: 12:00

ABSTRACT

To investigate whether people comprehend written music as they encounter it, we conducted two experiments in which expert pianists' eyes were tracked while they read and played single-line melodies. The melodies were either congruent or incongruent, with the incongruity involving a harmonic change of a single chord (e.g., non-resolution of an implied dominant seventh). Such incongruity led to rapid disruption in participants' reading, particularly in terms of regressions from the target bar, and also on measures of subsequent reading. Similar results occurred whether participants played with a metronome (Experiment 1) or without a metronome (Experiment 2), though the presence of a metronome led to rapid pupil dilation in the incongruent melodies. These results suggest that pianists rapidly interpret written music in relation to auditory music, just as readers interpret written language in relation to speech.

1. INTRODUCTION

One reason to expect incremental processing of written music is that written language is processed incrementally, almost word-by-word. Eye-tracking studies have shown that various types of linguistic anomalies cause disruption at early processing stages, with misspelt, ungrammatical, or implausible words each showing rapid effects (see Rayner, 1998 for a review). The question of this paper therefore is whether similar disruption occurs during musical comprehension, taking up the challenge proposed by Madell and Hebert (2008) of focusing on fine-grained aspects of music processing while reading notation. We report two experiments that investigate the effects of anomaly on music reading. Trained pianists read short musical scores that did or did not contain a contextually anomalous bar (as judged by a pretest) and then played the score on a piano. Our primary interest was in whether and how participants would be disrupted by the anomaly.

Incremental processing of spoken language was first demonstrated by Marslen-Wilson (1973), who showed that listeners are able to shadow normal prose accurately at a delay as short at 300ms, with good memory of the shadowed material and few structural errors. Since syntactic and semantic linguistic structures occur over far greater latencies, such structures clearly do not constrain processing. The importance of syntactic and semantic structure during written language processing has been addressed through studies of eye movements, in which readers have been found to experience rapid difficulty when a text becomes anomalous. When silently reading individual words, difficulties arise both when anomalies are non-contextual, such as in the case of misspellings (Underwood, Bloomfield, & Clews, 1988), and when anomalies are dependent on context, such as the case of semantic or syntactic implausibility (Rayner, Warren, Juhasz, & Liversedge, 2004; Stewart, Pickering, & Sturt, 2004; Braze, Shankweiler, Ni, & Palumbo, 2002).

In the case of contextual anomaly, similar eye movement effects have been noted for both syntax and semantics. Semantic anomaly was explored by Rayner et al. (2004) and involved participants reading plausible and implausible sentences, for example: 'John used a knife to chop the large carrots for dinner', and 'John used the pump to inflate the large carrots for dinner' respectively. First-pass time was significantly longer for the target region ('carrots') in the implausible condition than the plausible condition. Additionally, a significantly higher rate of regression from the posttarget region ('for dinner') was found for the implausible condition in comparison to the plausible condition. Other studies such as Braze et al. (2002) explored syntactic anomaly, also finding more regressions and longer first-pass reading time in the incongruent condition.

If music reading occurs incrementally, an expected effect would be that a musical anomaly would lead to rapid disruption in a standard reading situation. A musical analogy of a contextual anomaly in language would be the presence of a 'wrong note', or 'inappropriate chord progression' in an otherwise conventional piece. These notes or chords are not wrong in themselves, but are out of place in context. Tonal music is built on the tonal scale, which defines a hierarchy of stability and norms regarding acceptable progressions. This therefore forms one type of context against which incongruity can be judged, and since harmonic syntax is based on this scale, it has been used to replace contextual constraint in a number of studies of musical anomaly (see Patel, 2003, for a review). Although it is unclear as to whether this type of anomaly is syntactic or semantic in type, language studies have found similar effects occurring for syntactic and semantic anomalies making the relevance of the syntactic/semantic division in music inconsequential (see Patel, 2003; Koelsch, 2005; 2011 for various views).

Over the past 20 years eye movement studies of music reading have focused on large-grain measures such as the perceptual span and the eye-hand span. There has been little analysis of how musical factors influence reading. One previous study explored the effect of anomaly within music reading (Ahken, Comeau, Hébert, & Balasubramaniam, 2012). Pianists' eye movements were tracked while they read irregular and regular versions of music and language stimuli. The study additionally examined the effect of musical presentation-type, with some musical stimuli being presented with key signatures, and some with accidentals. In the language and key-signature music conditions, the proportion of fixations was larger for the incongruent than the congruent stimuli and the mean fixation duration was longer for the incongruent than the congruent stimuli. However, the music accidentals conditions did not display these effects. Several methodological issues raise concern in this study, and the dependence of congruency effects on musical presentation style suggests that the visual presentation of the music affected pianists' performance, rather than the musical progressions. In the current experiments, we focused on the effects of tonal expectation generated through implied harmony of a single-line melody, avoiding potential confounds by keeping the target bar constant across conditions, and excluding the use of additional accidental symbols in the target. To investigate the effect of timing constraints we additionally examined differences related to metronome use.

2. EXPERIMENT 1

2.1. Method

Participants 24 pianists were included in the analysis.

Stimuli All melodies were composed by the first author, and were 8-bars long to be played in the right hand of the piano. Accidentals were not used in target bars, and expectation was set up through use of either a sequence, or an implied harmonic progression requiring resolution. Each item comprised two congruous and two incongruous stimuli.

Procedure Eye movements were recorded using the SR EyeLink 1000 remote eye-tracker, sampling the left eye at 500Hz. Corneal and lens reflections were adjusted individually for each participant, with calibration occurring at the start of each block and any point that misalignment was evident.

Participants were asked to play in time with a metronome. The metronome speed was chosen so that quavers lasted 300ms. Participants were asked to begin playing as soon as possible, and keep to the beat of the metronome. They were told to continue without correction if a mistake was made. Sessions lasted approximately an hour.

Data Analysis We excluded participants' data if at least 50% of their experimental trials contained one or more pitch errors. We also excluded participants' data if at least 25% of experimental trials included fixations beyond the third bar of music before 2 seconds had elapsed, as this suggested that participants were scanning the score and preprocessing the stimuli before playing.

2.2. Results

Stimuli were divided into four regions of interest for analysis: (a) the signatures, (b) the pretarget bar, (c) the target bar, and (d) the posttarget bar (see Figure 1). Paired t-tests were computed between the congruent and incongruent conditions both by participant (F_1) and by item (F_2), after outliers were removed.

The analysis software outputs a number of eye-movement measures that can be used to infer cognitive processing, including first-pass reading time, and regression measures (Rayner, 1998). First-pass reading time is the summed fixation duration from the beginning of the first fixation in a region to the end of the last fixation in the region before the eye leaves that region, and is a measure of early processing activity. Regression measures include rate of regression out of a region, which is the proportion of trials in which the region is exited to the left prior to leaving the region to the right, and rate of regression in, which is the proportion of trials in which the region is entered from a region further to the right. Both first-pass reading time and regression rate out of a region are assumed to reflect early processing stages.

In the target bar there was a significantly higher rate of regression out of the target regions in the incongruent melodies (congruent M=0.086, incongruent M=0.184, F_1 & F_2 $p < 0.005$). This combined with a marginally significant higher rate of regressions back into the target region for the incongruent melodies. In the posttarget region first-pass time was significantly shorter in the incongruent condition. Exploratory analysis of pupillometry, shown to reflect processing difficulty in tasks involving high cognitive load (Piquado, Isaacowitz, & Wingfield, 2010), showed strong effects of congruency. Average pupil size was significantly greater in the target bar of the incongruent condition, an effect that spilled into the posttarget bar.

2.3. Discussion

Our main finding in this experiment was the significant effect of congruency on regression rates. However, contrary to the hypotheses drawn from the language comparison, we did not find any effects of congruency on first-pass fixation time measures in the target bar. In order to explore whether the metronome forced people to move on more quickly than they would naturally, hence disguising fixation time effects, we ran a second experiment without any tempo constraints. This additionally allowed us to replicate and generalise our findings in a typical reading situation, when pianists are free to manipulate their reading speed depending on the stimulus.

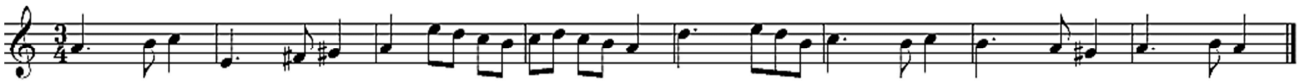


Figure 1: Example congruent melody. Signature, pretarget, target, and posttarget regions (highlighted in order).

3. EXPERIMENT 2

3.1. Method

In this experiment we used the same design as in Experiment 1, but without a metronome. If a participant asked what speed to play, they were told to play at a speed that felt natural. They were asked to begin playing as soon as possible after presentation of the notation, and were told to continue without correction if a mistake was made. 24 pianists were included in the analysis.

3.2. Results

Again, in the target region differences between conditions primarily revealed themselves in measures of regressions (congruent M=0.105, incongruent M=0.1.70, F_1 & F_2 $p < 0.06$). Significant differences were found in both the rate of regressions out of the target bar, and rate of regressions into the target bar. Regarding the pupillometry analysis, no effect of congruency was evident in the target bar.

3.3. Discussion

In this experiment, congruency was found to have the predicted effects on regressions within the target bar. Specifically, the target bar in the incongruent melodies elicited significantly more regressions to earlier portions of the notation. Regressions from the target bar may therefore have returned to the signatures region, which would be understandable as perhaps the most basic check of preceding context: the musical key. Significantly more regressions were also found to enter the target bar in the incongruent melodies, suggesting a subsequent checking mechanism.

Finally, a weak effect of congruency was found in maximum pupil dilation during the posttarget bar, but this was not consistent across pupillometry measures. A possible explanation for this could be that timing was not enforced by a metronome, which may have lowered the perceived difficulty of the task. However, differences in pupil size effects should be interpreted with caution, as people were slightly closer to the screen in one condition than another. For further analyses, see Hadley, Sturt and Pickering (2014).

4. GENERAL DISCUSSION

Several key results replicated across experiments 1 and 2, the most striking being that implied harmonic congruence had a significant and early effect on regression behaviour during music reading. In both studies, pianists made significantly more regressions from the target bar in the incongruent condition than the congruent condition. As the two conditions differed only in the intervallic relationship between the target bar and the surrounding context, this difference can be strongly linked to the harmony underlying the melodies, which either reflected a common progression, or an unusual progression. This provides clear evidence of incremental processing during music reading, which is particularly striking given that in the pilot participants showed only subtle differences in how 'natural' they judged the congruent and incongruent melodies. Using a 7-point scale, the average difference was only 0.94.

Highly significant effects were found in the pupillometry data of Experiment 1 (with metronome), but not in Experiment 2 (no metronome). Pupil dilation is a recognised measure of cognitive load, and supports the idea that incongruent melodies were more difficult to process than congruent melodies. In Experiment 1 this effect was strong, occurring across pupillometry measures in both the target and posttarget bars. In Experiment 2 however, these effects were not replicated and we suggest that this is due to the lack of metronome. Playing in time with an external beat is stressful, and makes errors obvious as they cannot be camouflaged or compensated for temporally. When playing freely, musicians encountering difficulty can slow down and use other expressive techniques to compensate, giving themselves more time to evaluate the upcoming passage. The lack of this possibility in the metronome condition may have led to the difficulty being revealed in pupillometry as opposed to performance timing.

Our findings from music reveal several similarities between reading in music and in language. Through significant differences in regression rate between congruent and incongruent melodies this study has shown that music is processed incrementally during reading, similarly to language. We have additionally shown that musicians implement complex checking procedures throughout performance, often returning to the signature regions during periods of uncertainty. An interesting possibility is that musical incongruence is reflected not only in regression measures but, when temporally constrained, in pupil dilation. This is the first time that either regression rate or pupil dilation has been linked to musical congruence during reading. Our research also indicated several differences between language and music reading. For example, we did not find effects in first-pass reading time. Further research investigating music reading and language reading within the same participants could elucidate such differences. Following these findings we suggest that music reading could provide a valuable new perspective from which to approach the music-language debate, and that the use of eye-tracking in music could provide insights not only into musical expectation but also factors underlying performance expression.

5. REFERENCES

Ahken, S., Comeau, G., Hébert, S., & Balasubramaniam, R. (2012). Eye movement patterns during the processing of musical and linguistic syntactic incongruities. Psychomusicology: Music, Mind and Brain, 22*(1), 18–25.*

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. (2002). Readers' eye movements distinguish anomalies of form and content. Journal of Psycholinguistic Research, 31*(1), 25-44.*

Hadley, L.V., Sturt P., & Pickering, M. (In manuscript). The Effects of Anomaly on Music Reading: Evidence from Eye Movements.

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. Current opinion in neurobiology, 15*(2), 207–12.*

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. Physics of life reviews, 8*(2), 89–105.*

Madell, J., & Hébert, S. (2008). Eye Movements and Music Reading: Where Do We Look Next? Music Perception: An Interdisciplinary Journal, 26*(2), 157–170.*

Marslen-Wilson, W. (1973). Linguistic Structure and Speech Shadowing at Very Short Latencies. Nature, 244*(5417), 522–523.*

Patel, A. D. (2003). Language, music, syntax and the brain. Nature neuroscience, 6*(7), 674–81.*

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. Psychophysiology, 47*(3), 560-569.*

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological bulletin, 124*(3), 372–422.*

Rayner, Keith, Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. Journal of experimental psychology: Learning, memory, and cognition, 30*(6), 1290–301.*

Stewart, A. J., Pickering, M. J., & Sturt, P. (2004). Using eye movements during reading as an implicit measure of the acceptability of brand extensions. Applied Cognitive Psychology, 18*(6), 697–709.*

Underwood, G., Bloomfield, R., & Clews, S. (1988). Information influences the pattern of eye fixations during sentence comprehension. Perception, 17*(2), 267–78.*

Session 7

[7A] Symposium 2

SEOUL 14:00- 16:00 Thursday 7 Aug 2014

MUSICAL MOVEMENT: EFFECTS OF CUEING AND FEEDBACK

Discussant: Peter Keller, University Of Western Sydney, Australia

Introduction

Joint musical performance is a highly multifaceted process, the elements of which are often investigated in isolation. The entrainment of motor output to an auditory stimulus, which happens in music performance but also in dance or cued movement rehabilitation, is in itself a rich research topic. Concurrently receiving auditory feedback (or sonification) of one's own movements provides additional information about the performance, which in turn can finetune movement precision, and may also support motor learning. In the social context of joint music making, interpersonal coordination necessitates anticipating others' actions, for which visual processing may also be included as entrainment information. Although these different processes happen simultaneously in joint music performance, their influence on motor output and motor learning have traditionally been studied separately. By bringing together scientific findings of movement cueing and feedback from a range of research groups, this symposium aims to integrate the ongoing discussion on entrainment and sonification, and highlight their applicability to musical practice, with

applications in music pedagogy as well as movement rehabilitation. As different measurement methods inherently address different aspects of this issue, a range of experimental approaches and subject populations are brought together in the symposium. Brain imaging results of healthy volunteers are presented, showing brain-level differences in the motor network for movement entrainment for differing cues, including a first investigation of moving to imagined music. Taking a more naturalistic approach, two motion capture studies are presented; one with results from expert drummers, demonstrating the effects of training on synchronization abilities, and the other demonstrating that in musician dyads, facing each other improves synchronization. Finally, applied clinical findings elucidate the impact of auditory feedback on motor learning for both healthy volunteers and stroke patients, indicating that optimal learning conditions do not necessarily generalize to patient populations, thus offering a practical approach to health applications. Through this symposium, we aim to stimulate discussion within the music cognition research community that reaches beyond the isolated aspects of musical experiences to a holistic approach to music practice and its clinical implications.

fMRI MEASURES OF CUED MOVEMENT

Rebecca S. Schaefer[1], Alexa M. Morcom[2], Neil Roberts[3], Katie Overy[4]
[1]SAGE Center for the Study of the Mind, University of California, Santa Barbara, USA. [2]School of Philosophy, Psychology & Language Sciences, University of Edinburgh, UK. [3]Clinical Research Imaging Centre (CRIC), The Queen's Medical Research Institute (QMRI), University of Edinburgh, UK. [4]Institute for Music in Human and Social Development, Reid School of Music, University of Edinburgh, UK.
Time: 14:00

Moving to music is widespread; auditory rhythms are known to induce as well as guide movement. However, the brain mechanisms of this guiding of movement are poorly understood, as studies of auditorily entrained movement show inconsistent results with regards to the involvement of cerebellar, premotor and basal ganglia areas. Recent investigations also identify imagined music as a possible movement cue, the effects of which might be mediated by the same neural circuits as for perceived music. We investigated how auditory cueing affects movement-related brain activity, using metronome cues that are common in movement rehabilitation settings as well as a complex, ecologically valid music stimulus. Additionally, we conducted the first investigation of moving to imagined music. Two functional magnetic resonance imaging (fMRI) experiments were performed with 17 nonmusician volunteers, performing simple wrist flexions. First, basic cueing was investigated, contrasting a metronome with uncued movement. Next, the effect of cueing was evaluated for music and imagined music, again contrasted to uncued movement. Each cueing condition (metronome, music, imagined music) was compared to self-paced movement within a region of interest mask for motor areas (BA4, BA6, basal ganglia, motor nuclei of the thalamus (ventral anterior nucleus & ventrolateral nucleus) and cerebellum). We also measured each cueing condition with motion capture in a separate behavioural experiment, revealing no significant differences between conditions in the number of flexions, range or speed of movement. The fMRI data revealed that compared with uncued movement, metronome cues led to activation in left ventral premotor and middle frontal premotor areas. Using music as a cue however revealed engagement of left cerebellar lobule VI as compared to self-paced movement. Conversely, movement cued by imagined music led to activity in the right pallidum. Whereas the cortical activations found for metronome cueing may be related to finer motor control, the cerebellar activation for music is interpreted as related to auditorymotor connectivity. The pallidum activation during imagery may reflect the increased effort of the task of imagining music. These neural activation differences, given the same movements, suggest that the type of movement cue can lead to specific motor network activation patterns. The results suggest that in clinical settings, different auditory cues may be suitable to cue movement in different patient groups.

TEMPO EFFECTS ON MOVEMENT KINEMATICS: A MARKER OF EXPERTISE IN PROFESSIONAL DRUMMERS

Nicolas Farrugia[1,2], Simone Dalla Bella [2,3]
[1]Goldsmiths, University of London, London, United Kingdom. [2]Movement to Health Laboratory (M2H), EuroMov, Montpellier-1 University, Montpellier,
France. [3]Institut Universitaire de France (IUF), Paris, France
Time: 14:30

Skilled musicians are used to play with the beat at different tempos. The way their movements adjust to tempo for achieving high spatial and temporal accuracy is likely to reflect how audiomotor coupling shapes experts' movements. Indeed, motor strategies are built through training, and are commonly used as pedagogical aims (e.g. keeping fingers close to the keyboard to perform fast melodies). In particular, effects of tempo on anticipatory movements are reported in previous studies on piano and clarinet performance, e.g. larger movement amplitude with increasing tempo. In the present study, we assessed whether anticipatory movements in drumming vary as a function of musical training in synchronization. We were interested in (1) studying the effect of tempo on anticipatory movement and synchronization accuracy and variability, (2) relating anticipatory movement to temporal accuracy, and finally (3) explaining observed individual differences in movement patterns and synchronization by the amount of musical training. Eight professional drummers with varying degrees of musical training participated in this study. Participants synchronized to an isochronous sequence of tones presented at different tempi, spanning from 60 to 200 BPM. The movement of their right arm and stick was captured by a Vicon system during performance using passive reflective markers. The stick trajectory was analyzed in terms of maximum distance to the pad (movement amplitude), and time when this maximum is reached (movement anticipation). Results show that tempo modulated both synchronization accuracy and movement kinematics. Participants were less accurate and more variable at the fast tempi than at the slower tempi. They raised less the arm and anticipated more the stroke at fast than at slow tempi. Interestingly, both amplitude and anticipation time predicted synchronization accuracy at all tempi, while variability was mostly related to anticipation time. Finally, several measures of musical training could explain observed individual differences in synchronization as well as in movement patterns. Both timing and amplitude of anticipatory movements in drumming contribute to the optimization of sensorimotor performance. The underlying audiomotor coupling mechanism is developed progressively via training. These results pave the ground to the use of enhanced feedback (e.g., sonification) for improving sensorimotor training.

EFFECT OF TEMPO AND VISION ON INTERPERSONAL COORDINATIONG OF TIME

Marc R. Thompson[1], Tommi Himberg[2], Justin London [3], Petri Toiviainen[1]
[1]Finnish Centre in Interdisciplinary Music Research, University of Jyväskylä, Finland. [2]Brain Research Unit, O.V. Lounasmaa Laboratory, Aalto University. Espoo, Finland. [3]Department of Music, Carleton College, Northfield MN USA

Time: 15:00

Interpersonal coordination within a dyadic musical performance requires that the two musicians share a similar mental model of the music's timing structure. In addition to performing with agreeing inter-onset-intervals (IOIs), matched mental models can be observed through corporeal articulations (i.e. synchronous body sway, mimicked or complementary gestures). While performers may depart from the stringent pulse set by a metronome at the onset of a performance, interpersonal coordination persists through continual auditory and visual feedback. Keller and Appel (2010) tracked the movements of pianist dyads and found that interpersonal synchrony is enhanced when musicians perform in view of one another. This study furthers this line of research by 1) using instruments that allow for a greater range of motion (i.e. strings and woodwinds) and 2) using full-body motion capture for kinematic analysis. Our aim was to examine the effect of tempo on interpersonal coordination within a musical dyad. Musician dyads performed three unfamiliar collaborative musical sequences in facing vs. non-facing conditions. Our hypotheses were that interpersonal coordination would be weakened in the non-facing conditions, and that synchronization would be affected by both very slow and very fast tempi. Three musical sequences consisting of single or pairs of alternating tones were performed at four different tempi (60 BPM, 90 BPM, 120 BPM & 150 BPM) in facing and non-facing conditions (3x4x2 design). Each tempo was indicated with a eight-beat metronome count-off, the metronome stopped when the musicians began playing. The sequences were of increasing complexity, ranging from alternating repeated tones (to simulate a tapping paradigm) to interleaved two-note scale patterns, e.g. "do-re," "re-mi," "mi-fa," (etc.). We collected data from six musician dyads (four violins, one clarinet, and one flute dyad). Both the audio and movement patterns are recorded and analyzed. After coding the IOIs in each performed sequence, we performed cross-correlations within each dyadic performance. The maximum coefficient was consistently found at lag -1. We then correlated these coefficients with mean and standard deviation IOIs from each performance for the facing and non-facing conditions. Results showed general a trend for musicians to speed up from the initial BPM in the facing condition, versus maintaining more steady tempo, or slowing down in the non-facing condition. However, they remained more synchronized in the facing versus the non-facing condition. The results of the facing condition's effect on corporeal entrainment within the dyads will be discussed during our presentation.

THE BRAIN USES AUDITORY FEEDBACK FOR MOTOR LEARNING OF TIMING IN HEALTH AND STROKE POPULATION

Floris T. Van Vugt[1,2], Eckart Altenmueller[2], Barbara Tillmann[1]
[1]Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, CNRS-UMR 5292, INSERM U1028, University Lyon-1, Lyon, France. [2]Institute of Music Physiology and Musicians' Medicine, University of Music, Drama and Media, Hannover, Germany
Time: 15:30

Sensorimotor synchronisation, participants perform movements in time to predictable external events (e.g., a metronome). In this paradigm, the external events do not depend on the movement. However, when humans play a musical instrument, the produced sounds depend on their movement. This would in principle allow the brain to compare the sound output to a template and make trial- by-trial corrections to approach the template, resulting in motor learning. We were interested to study how sounds can function as a feedback signal in this way and improve the brain's motor control. We performed two experimental studies to test this idea. In a first study, we tested whether auditory feedback is used by healthy participants in motor learning. In a second study we tested whether auditory feedback can also benefit patients undergoing music-supported motor rehabilitation after stroke. In the first study, we investigated 36 healthy non-musician participants who learned to tap a sequence of keystrokes as regularly as possible in time. Participants were divided into three groups. In the sound group, a tone was presented immediately every time the participants pressed a key. In the mute group, no keystroke-triggered sound was presented. The jitter group received sounds with a random delay of 10-190 msec after keystroke, disrupting the time-lock between movement and sound. In the second study, 34 patients (less than 6 months post-stroke) were selected to form a representative sample of the clinical population at the rehabilitation clinic. Patients had moderate movement impairment but residual motor function. In our study, patients learned to play finger exercises and simple songs on a keyboard that either emitted sounds immediately (normal group) or with a jittered delay. In the first study, participants in the sound group were able to improve in tapping regularity whereas the mute and jitter group did not improve. In the patient study, surprisingly, the jittered delay group showed greater improvement than the normal group in clinical transfer measurements. In sum, we show that auditory feedback is used to learn motor control of regularity. Further, the time-lock between movement and sound is critical for this learning. Counter-intuitively, disruption of this time-lock may be beneficial for stroke populations.

[7B] Cognitive Neuroscience of Music 3

KYOTO 14:00- 16:00 Thursday 7 Aug 2014

NEURAL TRACKING OF MUSICAL MOTIVES REVEALED BY A COMBINATION OF FMRI AND MUSIC INFORMATION RETRIEVAL TECHNIQUES

Tom Collins[1], Daniel A. Abrams[2], Rohan Chandra[2], Christina Young[2], Andreas Arzt[1], Vinod Menon[2]
[1]Department of Computational Perception, Johannes Kepler University Linz, Austria, [2]School of Medicine, Stanford University, USA
Time: 14:00

1. Background: Music's capacity to affect people profoundly is based in large part on the formation and activation of musical memories. fMRI studies on musical memory reveal differing and diverse brain activity for familiar versus unfamiliar stimuli. It is unclear, however, whether this is due solely to the familiar-unfamiliar variable, or to other uncontrolled differences in pitch and rhythmic material of stimuli. The field of music information retrieval (MIR) has produced myriad algorithms for retrieving the location(s) of a musical motive in a given piece, as well as the strength of matches. This pattern matching technology might enable neuroscientists to better isolate the familiarity variable and its corresponding neural response, by studying activity associated with specific motive occurrences.

2. Aim: The aim of the study is to refine our understanding of musical memory, by identifying brain structures that track occurrences of a motive throughout an extended and naturalistic music listening experience.

3. Method: Brain activity was measured using fMRI while non-musician adults listened to the first movement of Ludwig van Beethoven's (1770-1827) Fifth Symphony. All listeners reported familiarity with the piece prior to scanning. To identify neural structures that are sensitive to the

relative strength of the famous opening four-note motive, we performed a regression analysis of the fMRI time series data collected in every voxel of the brain in each of our 17 listeners, on a repetition strength time series generated by an MIR pattern matching algorithm.

4. Results: Group results from the regression analysis indicate that two brain regions, the left-hemisphere hippocampus and bilateral medial dorsal nucleus of the thalamus, are associated with the relative strength of the four-note motive throughout the symphonic movement. Analogous analysis for a less well known motive (Violin I, measures 63-66) did not reveal significant responses in these regions, suggesting the findings are linked specifically with musical familiarity rather than a general stimulus response.

5. Conclusion: By combining fMRI and MIR techniques, we have isolated more precisely the neurological basis of musical familiarity. We have identified for the first time the neural structures that appear to track occurrences of a musical motive during music listening.

UNDERSTANDING THE PERCEPTION OF MUSIC FROM THE PERSPECTIVE OF OSCILLATORY ACTIVITY WITHIN AND BETWEEN BRAIN NETWORKS

Frank Russo[1]
[1]*Ryerson University, Canada*
Time: 14:30

1. Background: Synchronization of action appears to underlie a wide range of musical activity. Increasingly, researchers are examining the extent to which these activities may rely on the synchronization of neural oscillations.

2. Aims: In this talk, I will review research on music from the perspective of synchronization. In particular, I will discuss the evidence for magneto- and electroencephalography (M/EEG) synchronization within and between neural networks that may underlie the perception of music. More generally, the review will attempt to connect recent research in the neural dynamics of music to the broader trend in cognitive neuroscience of understanding human cognition through the lens of oscillatory activity in brain networks.

3. Main contribution: The relations between musical perceptions and oscillatory activity will be considered in subcortical and cortical brain networks. After a discussion of general principles underlying synchronization, I will consider research concerning the frequency-following response derived from M/EEG recordings of brainstem activity and how it may be used to predict precision in frequency discriminations. Second, I will consider research concerning beta-band synchronization to pulse in auditory cortical networks and how it may relate to expectancy and sensitivity to rhythmic change. Third, I will examine mu desynchronization that has been localized in the sensorimotor cortex and that appears to relate to the motor simulation of emotional expression. Finally, I will evaluate the extent to which these various types of synchronization may be related and how they may interact so as to affect musical perception.

4. Implications: M/EEG methods have millisecond accuracy and have the potential to elucidate the manner in which perception, expectancy and emotion vary across individuals and over time. When combined with advanced source localization methods, this approach can be used to develop a comprehensive understanding of the oscillatory dynamics of music. A better understanding of these dynamics may eventually inform approaches to the development of assistive technology (e.g., hearing aids) and music-based neurological therapies.

TOWARD AN OBJECTIVE MEASURE OF LISTENER ENGAGEMENT WITH NATURAL MUSIC USING INTER-SUBJECT EEG CORRELATION

Blair Kaneshiro[1]*, Jacek Dmochowski*[1]*, Anthony Norcia*[1]*, Jonathan Berger*[1]
[1]*Stanford University, USA*
Time: 15:00

ABSTRACT

This study extends existing research on inter-subject correlations (ISCs) of brain responses as a measure of engagement, with a focus on the relationship between the structural coherence of a musical stream and the listener's degree of engagement with it. EEG was recorded while subjects listened to naturalistic music (popular Hindi songs) presented in original versions and in temporally disrupted, phase-scrambled conditions. ISCs were computed from the EEG data using Reliable Components Analysis. Overall, original versions of songs yielded significantly higher ISCs than phase-scrambled versions, and were also rated as more pleasant, well ordered, and interesting by subjects. The most reliable spatial component extracted from responses to the original songs concurs with past EEG findings involving naturalistic music. The time course of the ISCs is resolved at a musically relevant time scale. The sum of our findings suggests that ISCs show promise toward finding time-critical measures of engagement and attention in typically noisy EEG signals, and, specifically, as a means to find correlations between structural features of music and brain responses in listeners. We discuss possible links between heightened ISCs and regions of musical interest, and implications for future research.

1. INTRODUCTION

Listening to music is generally regarded to constitute a pleasurable endeavor. The sense of pleasure and the degree to which a listener engages with the music are interdependent. Varying degrees of engagement can occur whether the music is the focus of attention (for example when heard in a concert hall), or is the accompaniment to other activities such as socialization or work. Even when music is heard passively as background to another task, arousal and attention levels can fluctuate in response to varying musical and acoustical features. Many of us are familiar with the feeling of engagement with music, whether hearing a song for the first time or the thirtieth. But how is engagement represented in the brain?

The use of inter-subject correlations (ISCs) as a measure of engagement stems from the reasoning that neural activity at any given time comprises exogenous (stimulus-driven) and endogenous (internally generated) components. When engagement with a stimulus increases, so too does the exogenous component of the brain response as peripheral processing decreases; therefore, responses become more correlated across subjects.

To date, ISCs have been examined in a variety of fMRI studies employing naturalistic stimuli including video (Hasson et al., 2004), speech and non-speech sounds (Honey et al., 2012; Boldt et al., 2013), and music (Abrams et al., 2013). In a recent EEG study by Dmochowski et al. (2012), a novel signal-decomposition method was devised for extracting maximally correlated components from neural responses to videos. This

method was then used to quantify viewer engagement on a finer time scale, which allowed periods of heightened engagement to be traced back to contextually salient scenes in the videos.

In the present study, we combine the methodology of Dmochowski et al. (2012) and Abrams et al. (2013), using ISCs of EEG responses as a measure of engagement with naturalistic music. We do this by identifying correspondences among brain responses to naturalistic music that is presented in original and control conditions. Behavioral ratings of the stimuli serve as a point of comparison for the EEG results. We employ 'scramble' paradigms similar to those used in past studies (Abrams et al., 2013; Dmochowski et al., 2012; Levitin & Menon, 2003) to disrupt the temporal coherence of the musical stimuli, with a current focus on very small-scale temporal manipulations achieved through phase scrambling. While such low-level temporal units do not necessarily constitute meaningful structural units in a musical sense, we begin our line of research at this level, as a basis for building up to larger structural components of music (e.g., measures, phrases, and song parts).

Our aim in the current study is to validate ISCs of EEG-recorded brain responses as a reliable measure of engagement. Our broader goal aims to establish an objective measure of listener engagement with naturalistic music, which could provide a useful supplement to existing physiological and self-report measures. EEG proves to be a useful modality for this, as its temporal resolution is well matched to that of music, meaning that there is potential to draw connections between heightened ISCs computed over small time windows and corresponding features of the driving stimulus. At the same time, the ISC source-selection technique used here requires a listener to hear each stimulus only once (provided the stimuli are sufficiently long), and does not rely upon event-related averaging (Ben-Yakov et al., 2012). This allows us to include a greater variety of stimuli, and more importantly, captures the real-world experience of engaging with a musical excerpt in a single listen.

2. METHODS

2.1. Stimuli

Song selection. We used an unorthodox stimulus set derived from four songs from recent popular Hindi-language films. We did so because we sought songs in a (relatively) tonal idiom that have appealed to a massive audience and would presumably engage a listener, yet would be unfamiliar to our experiment subjects. While lyrics form an integral component of engagement with popular music, we wanted to avoid in the current study any effects of the semantic content of lyrics. We additionally concluded from a previous pilot study that non-English lyrics would be less unsettling to the subject in control conditions that noticeably disrupted the flow of lyrics.

Songs chosen for the study (summarized in Table 1) were released as singles from their respective films, or featured prominently in the films. All are sung in Hindi dialects with minimal English lyrics;21 include clear verse and chorus elements; use a steady beat throughout; and comprise regularly structured phrases. We acknowledge some differences in vocalization and instrumentation, but attempted to maximize the 'Westernness' of the song set, in style and instrumentation, given the above constraints.

Table 1. Songs used in the study.

	Song Name	Film	Year	Length
Song 1	Ainvayi Ainvayi	*Band Baaja Baaraat*	2010	4:27
Song 2	Daaru Desi	*Cocktail*	2012	4:30
Song 3	Haule Haule	*Rab Ne Bana Di Jodi*	2008	4:24
Song 4	Malang	*Dhoom 3*	2013	4:33

Control stimuli. The phase-scrambled control stimuli were created as described in Abrams et al. (2013), by randomizing the phase response at each frequency bin in a song's Fourier transform to a value between 0 and 2π, then transforming the signal back to the time domain. This manipulation effectively washes out the temporal structure of a song while preserving its magnitude spectrum. In addition—while we do not cover these results in the present study—beat-shuffled versions of the songs were created using freely available beat-tracking code (Ellis, 2007). Beat onset times were detected in the audio and used as segmentation points for permuting the beats of the songs. This manipulation thus results in a stimulus that retains the steady beat of the original song, but does not provide a coherent narrative for the listener in terms of phrase structure or melodic/harmonic continuity. Both types of control stimuli were created using Matlab.

2.2. Participants and Procedure

Twelve right-handed subjects aged 19-38 (mean age 28.17 years; 2 female) participated in the experiment.22 Formal musical training ranged from 0-18 years (mean 9.63 years), though all subjects either had formal training or had taught themselves to play an instrument. Participants were unfamiliar with the songs used in the study, and no participants spoke Hindi, listened to Hindi music, or watched Hindi films. All participants listened at least occasionally to popular music in English.

Each stimulus was presented once in its entirety, in random order. Stimuli were delivered at a comfortable listening level through magnetically shielded Genelec 1030A speakers while subjects were seated in a darkened, electrically and acoustically shielded booth. Subjects were instructed to attend to the stimuli but performed no behavioral task while audio was playing. Following each stimulus, subjects answered the following questions on a scale of 1-9:

1. How pleasant was the excerpt?
2. How well-ordered was the excerpt?
3. How much of the excerpt was interesting?

Stimulus delivery and collection of behavioral responses was performed using Neurobehavioral Systems Presentation software. 128-channel EEG (Electrical Geodesics, Inc.) was recorded at a sampling rate of 1 kHz, referenced to the vertex.

21 Songs included at most one- or two-word interjections or one-syllable word substitutions in English.
22 Data from a thirteenth subject were excluded from analysis due to gross artifacts.

2.3. Data Analysis

Preprocessing. Data were passband filtered between 0.3-50 Hz and downsampled by a factor of 4 using EGI's Net Station software; all subsequent analyses were performed in Matlab. Data were epoched, EOG channels were computed, and bad electrodes, as well as electrodes covering the face, were excluded from further analysis. Eye artifacts were removed using EEGLAB's extended Infomax ICA (Jung et al., 1998, Delorme & Makeig, 2004). Following this, DC offset was subtracted from each channel, and data matrices were converted to average reference. Transient samples whose magnitude exceeded 4 standard deviations of its respective channel's mean power were set to NaN.

Inter-subject correlations. ISCs were computed using Reliable Components Analysis (RCA), as described in Dmochowski et al. (2012). This technique is similar to Principal Components Analysis (PCA) and Independent Components Analysis (ICA), differing in that it optimizes reliability of data records (as opposed to variance explained or statistical independence, respectively). Namely, given a set of space-time data records (i.e., one for each subject), the algorithm computes projections of the data exhibiting maximal ISC. In other words, the criterion being maximized is the sum of pairwise ISCs among all unique subject pairs. For the analysis, we computed the Reliable Components (RCs) separately for original and phase-scrambled tracks, pooling covariances from all songs/subjects in the computation.

For each song, we then projected the neural responses of all subjects onto the corresponding first RC and computed the mean (across unique subject pairs) ISC time course in time windows of 10 seconds, with a 2-second shift between successive windows. This yields a time-resolved measure of ISC for all songs. For each time window, we performed a Wilcoxon signed-rank test to compute the probability that the measured ISCs were drawn from a zero-median distribution. This yields a proportion of time windows that exhibited significant ISCs ($p<0.05$) throughout the song duration.

3. RESULTS

3.1. Behavioral Results

Subjects' behavioral ratings of the stimuli along the dimensions of Pleasantness, Order, and Interestingness are shown in Figure 1. Original versions received higher ratings than phase-scrambled versions along all three dimensions; a nonparametric test of statistical significance finds the difference in ratings between song versions to be significant for each dimension (Mann-Whitney U test, $p<10^{-14}$). We observe a number of outlier ratings of phase-scrambled stimuli, especially for Order; we believe this occurred because we deliberately formulated that prompt in an ambiguous manner, and subjects could interpret the phase-scrambled stimuli as being either highly ordered (lacking variation over time) or not ordered at all (no discernible structure).

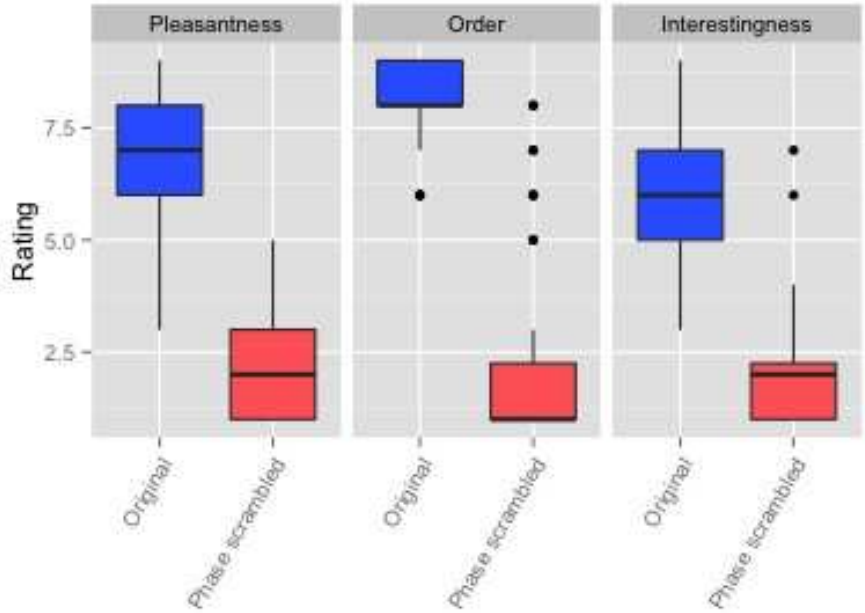


Figure 1. Subjects' behavioral ratings of original (blue) and phase-scrambled (red) versions of songs, along dimensions of Pleasantness (left), Order (middle), and Interestingness (right).

3.2. EEG Results

From the subject- and song-aggregated covariance matrices, we computed the first Reliable Component (RC1). Figure 2 shows the corresponding projections of the extracted activity—the so-called "forward-model" (Parra et al., 2005)—for original (left) and phase-scrambled (right) versions of the songs. RC1 for the original versions of the songs is marked by poles over frontocentral cortex and bilateral temporoparietal regions. Without the availability of detailed anatomical source inverses, it is difficult to speculate on the neural origins of the observed component. However, we do note that its topography is similar to the first Principal Component found in a published EEG study using naturalistic music (Schaefer et al., 2011), and is consistent with bilateral dipoles in temporal cortex. Meanwhile, the first RC stemming from data recorded during phase-scrambled stimuli exhibits a disparate topography with poles in the inferior occipital and occipitotemporal cortices, but lacking physiological plausibility.

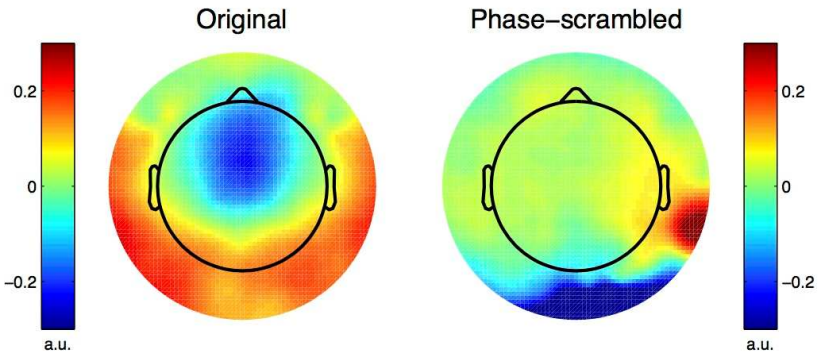


Figure 2. Forward-model projections of RC1 (most reliable component) extracted from responses to the original versions (left) and phase-scrambled versions (right) of stimuli.

The brain responses in the space of these components were used to compute the ISCs as a function of time for each stimulus, using 10-second temporal windows of data that advanced in 2-second increments. Results, grouped by song, are plotted as a function of time in Figure 3. Inspection of the plots suggests that phase-scrambled versions exhibit lower ISCs than original versions, for all songs.

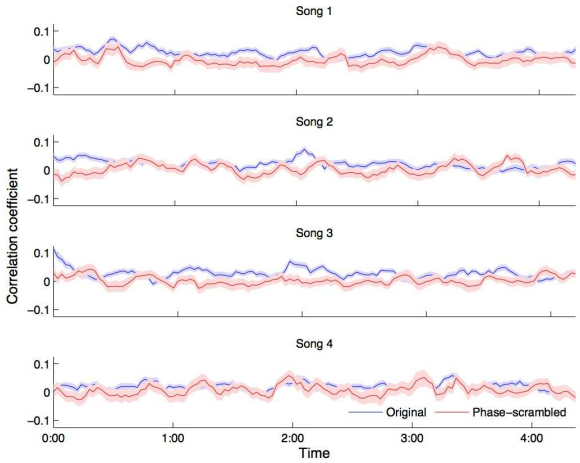


Figure 3. ISCs computed from the most reliable component (RC1) of original (blue) and phase-scrambled (red) versions of individual songs, plotted as a function of time. RC1 was computed separately for the original and phase-scrambled versions in order to maximize the ISCs of all EEG records pertaining to that condition.

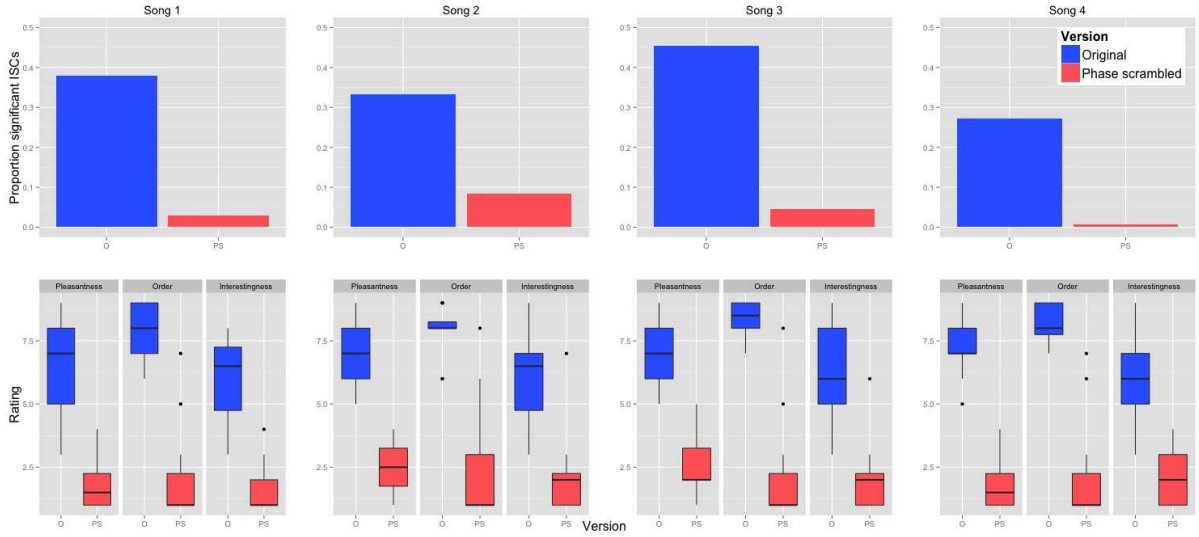


Figure 4. Top: Proportion of significant ISCs in the four songs, for original (blue) and phase-scrambled (red) versions. Bottom: Behavioral ratings of Pleasantness, Order, and Interestingness for each version of the songs.

Finally, we present the proportion of significant ISCs for each stimulus. The top portion of Figure 4 shows the proportion of significant ISCs for each stimulus condition of each song; corresponding behavioral ratings for each song are shown in the bottom portion of the figure. We see that the proportion of significant ICSs of original versions is always higher than the phase-scrambled versions, as are the behavioral ratings.

4. DISCUSSION

This work constitutes an initial step toward developing an objective, time-resolved measure of listener engagement with naturalistic music using ISCs and EEG. In this study, we used a single-listen paradigm to present popular yet novel songs in their entirety, both unaltered as well as temporally distorted through phase scrambling. The most reliable component obtained from EEG responses to original versions of songs using RCA appears to reflect bilateral dipoles in temporal cortex and is consistent with past EEG findings using naturalistic music. Subjects' behavioral ratings of the stimuli concur with ISC results: Original versions of songs were rated as more pleasant, well ordered, and interesting than phase-scrambled versions, while also driving higher ISCs in the brain responses.

We speculate that moments of high engagement, as measured by high temporally resolved ISCs (Figure 3), may correspond to particularly arousing points in the original songs. Indeed, our early observations do suggest a marked co-variation between peaks in the ISCs and structural demarcations of the songs correlating to various musical dimensions.23 If this proves to be the case, then it may be true that ISCs can serve to index listener engagement with naturalistic music with high temporal resolution, over just a single listen.

One question that arises is whether certain factors may draw attention without necessarily fostering engagement. For example, to what degree do predictable spikes in amplitude of an audio excerpt (as in many metrical accents) drive engagement? As we proceed with our research, we hope to explore and disentangle the interplay of arousal, attention, and more broadly, engagement. Dmochowski et al. (2012) characterize engagement as "emotionally laden attention." By incorporating subjective ratings with objective measures obtained from the EEG data, we hope to pursue this supposition in the context of listening to music.

Broadly speaking, using ISCs to find significant features in noisy EEG data offers a method to use ecologically valid, naturalistic stimuli—stimuli of greater duration and complexity—to study attention, arousal, and engagement. This study provides a referential point of departure for future research aimed at exploration of the effects of temporal organization and coherence on musical engagement. A next step is the analysis of higher-order control stimuli (e.g., re-orderings at the beat and phrase level). Furthermore, ISCs suggest a potential meaningful measure of engagement based upon other musical dimensions. Future studies will explore the effect of repeated hearings on engagement, effects of lyrics and language, and the effect of familiar versus novel songs as stimuli. One intriguing question is whether ISC-based engagement detection might be used predictively to assess likely audience reaction and behavior on a large scale (following forthcoming work by Dmochowski et al. (2014)). To this end, we intend to create a parallel research track using large-scale datasets tracking music engagement, and integrate statistical observations and models with inter-subject correlations in music engagement experiments.

5. ACKNOWLEDGMENT

This research was supported by the Wallenberg Network Initiative: Culture, Brain, and Learning. The authors wish to thank Sophia Laurenzi and Steven Losorelli for their assistance with data collection, and Shubhabrata Sengupta for his insights into characterizing the stimuli.

6. REFERENCES

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., & Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. European Journal of Neuroscience.

Ben-Yakov, A., Honey, C. J., Lerner, Y., & Hasson, U. (2012). Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. NeuroImage, 63, *501-506.*

Boldt, R., Malinen, S., Seppä, M., Tikka, P., Savolainen, P., Hari, R., & Carlson, S. (2013). Listening to an audio drama activates two processing networks, one for all sounds, another exclusively for speech. PLoS One, 8(5).

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134, *9-21.*

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention—a possible marker of engagement? Frontiers in human neuroscience, 6.

Dmochowski, J. P., Bezdek, M., Abelson, B., Johnson, J., Schumacher, E., and Parra, L. C. (to appear). Audience preferences are predicted by temporal reliability of neural processing. Nature Communications.

Ellis, D. P. W. (2007). Beat tracking by dynamic programming. Journal of New Music Research, 36(1), *51-60. http://labrosa.ee.columbia.edu/projects/beattrack/*

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. Science, 303(5664), *1634-1640.*

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: Neural responses shared across languages. The Journal of Neuroscience, 32(44), *15277-15283.*

Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (1998). Extended ICA removes artifacts from electroencephalographic recordings. Advances in Neural Information Processing Systems, 10, *894-900.*

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. NeuroImage, 20, *2142-2152.*

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. NeuroImage, 28, *326-341.*

Schaefer, R. S., Farquhar, J., Blokland, Y., Sadakata, M., & Desain, P. (2011). Name that tune: Decoding music from the listening brain. NeuroImage, 56, *843-849.*

EXPERT MUSICAL IMAGERY: AN FMRI INVESTIGATION

Kirsteen Davidson-Kelly[1], S. Hong[1], R.S. Schaefer[2], E. J. Van Beek[3], N. Roberts[3], K. Overy[1]

23 Examples can be found at http://ccrma.stanford.edu/groups/meri

[1]Institute for Music in Human and Social Development (imhsd), Edinburgh College of Art, University of Edinburgh, United Kingdom, [2]Sage Center for the Study of the Mind, University of California, Santa Barbara, United Kingdom, [3]Clinical Research Imaging Centre (cric), Queen's Medical Research Institute, University of Edinburgh, United Kingdom
Time: 15:30

1. Background: Musicians are expected to memorise a large and complex repertoire, but lack of clear memorisation pedagogy and consequent unsystematic and inefficient practice can contribute to performance anxiety and physical overuse syndromes. Evidence suggests that mental rehearsal can be used to reduce physical rehearsal and enhance memorisation, but it is not yet known exactly which techniques are most effective, or why. Previous neuroimaging research has found that expert musical imagery recruits similar neural networks to motor performance, but with interesting distinctions.

2. Aims: An fMRI study was designed to examine differences between musical imagery and motor performance, with the aim of elucidating reported advantages of mental rehearsal.

3. Method: This study of 14 expert pianists used novel, ecologically valid, bi-manual musical tasks to investigate the neural basis of musical imagery. A 2x2factorial design compared imagery and motor performance at two levels of musical complexity, while controlling for motor complexity. Participants memorised two short, novel, two-handed pieces of music immediately prior to scanning, following guided learning procedures. They were scanned (on a 3T Verio system at CRIC, Edinburgh) during imagery and simulated motor performance of the memorised pieces, without auditory feedback. All data were analysed using BrainVoyager QX.

4. Results: Activation in prefrontal regions (MFG, IFG) increased during imagery (versus motor performance) and when musical complexity increased. Furthermore, during imagery, professionals (n7) activated these regions significantly more than students (n7). Secondary auditory regions of the superior temporal gyrus (STG) were activated bilaterally during both imagery and motor performance. During motor performance, professionals activated left STG significantly more than students, and left STG activation increased when musical complexity increased.

5. Conclusions: For the first time, this study demonstrated that increasing musical complexity accentuated activation in prefrontal and superior temporal areas that have previously been identified as playing an important role in musical processing. Furthermore, these areas were increasingly activated by professional (compared with student) participants, which fits with evidence from other domains that imagery ability increases with experience. The fact that musical imagery preferentially activated prefrontal regions involved in working memory and attention processes supports the hypothesis that imagery rehearsal may particularly benefit musical memorisation.

[7C] Affections in Musical Performance

MONTREAL 14:00- 16:00 Thursday 7 Aug 2014

DEFINING EXPRESSIVENESS IN MUSIC: COGNITIVE AND CROSS-CULTURAL PERSPECTIVES

Emery Schubert[1], Dorottya Fabian[1], Renee Timmers[2]
[1]Unsw Australia, Australia, [2]University of Sheffield, United Kingdom
Time: 14:00

Music psychologists have explained expressiveness as deviation from the score or, in more sophisticated forms, deviation from a mental prototype. From a measurement perspective, this usually restricts expressiveness to deviations in tempo and dynamics. But recent researchers are revealing more nuanced ways of understanding expressiveness. Drawing from subdisciplines of philosophy, cognitive psychology and cross-cultural investigations, this paper examines contemporary approaches to propose a new conceptualisation of expressiveness in music. Doğantan-Dack warns us not to get trapped into defining expressiveness according to the measures we are using. For example spectrograms restrict our understanding of expressiveness to only the parameters that we are able to abstract from the spectrogram (which is not equivalent to the music). Cognitive perspectives help us to identify that expressiveness differentiates performances based on purely musical nuances, and these nuances must be based on mental representation, whether of a specific piece (e.g. a Chopin nocturne) or a general style (e.g. principles of improvisation in jazz). Without a memory of music repertoire, expressiveness must be a foreign concept. Recent cross-cultural research (e.g. by Marandola) reminds researchers that expressiveness in music is a highly Westernised construct but may nevertheless have relevance when investigating music of different cultures and styles: modernist, jazz, pop, and 'non-Western' musics. Traditional concept of expressiveness might not be explicitly relevant to these musicians, but understanding their performance through the lens of expressiveness is desirable in understanding music perception and action in general, even if confined to a Western research perspective. The conclusion of this paper, then, is to recommend that the conceptualization of expressiveness in music be based on three principle factors: (1) one performance must be judged as expressive against a style dependent norm, such as a prototypical or idealised performance, (2) expressiveness is a multifaceted phenomenon and the facet in question (structural, emotional and so forth) should be made explicit, and (3) these conceptualisations of expressiveness are to be made with the understanding that they may work in different ways when music outside the common practice period is considered.

THE EFFECTS OF CONTINUOUS SUBJECTIVE MUSIC EXPERIENCE RATINGS ON PHYSIOLOGICAL AND PSYCHOLOGICAL MEASURES OF ACTIVATION AND VALENCE

Hauke Egermann[1,2], Alexander Foerstel[1], Stephen Mcadams[2]
[1]Audio Communication Group, Technische Universität Berlin, Germany, [2]Schulich School of Music, McGill University, Canada
Time: 14:30

1. Background: Many studies use continuous rating devices to assess dynamic changes of experience during music listening. This creates an unnatural listening experience in which participants are required to constantly monitor both the music presented and their subjective responses concerning emotions, tension, or liking. However, to date, how these simultaneous ratings affect the dependent variables that are supposed to be measured remains unexplored.

2. Aims: Therefore, we conducted an experimental study in which we compared two randomly assigned groups of participants: one continuously rating induced emotions, one not. From both groups, measures of psychophysiological activation and retrospective ratings were taken. We

hypothesized that continuous ratings would lead to increased subjective and physiological activations during music listening due to participants' increased attentional focus.

3. Methods: 43 participants listened to 13 music excerpts from the classical repertoire (each approx. 30 sec. long) and retrospectively rated the intensity of emotional responses on a list of 28 items. Half of the participants additionally continuously rated their felt emotion in a two-dimensional emotion space of valence and arousal using an iPod-Touch interface. We also continuously recorded skin conductance, heart and respiration rate, and activation of facial corrugator and zygomaticus muscles (associated with frowning and smiling, respectively).

4. Results: Data analyses indicate that there were no differences in retrospective and physiological valence measures of induced emotions between participant groups. However, those that continuously rated their own emotions responded with significantly higher sympathetic nervous system activation indicated by higher skin conductance levels. This finding is furthermore corroborated by higher subjective arousal ratings in the retrospective questionnaire for the group with continuous ratings. Additionally, the participants in this group stated that they were less distracted by something other than the music during the task.

5. Conclusions: These results indicate that the task of continuously rating music experience influences music experience on its own, at least in terms of behavioral and physiological measures of arousal. We assume that this effect might be mediated through an increased attentional focus. Researchers employing this form of response paradigm need to consider this bias in interpreting their findings.

NEURAL SUBSTRATES OF PERCEIVED MUSICAL EMOTION DURING CONTINUOUS LISTENING

Petri Toiviainen[1], Elvira Brattico[2,3]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland, [2]Brain & Mind Laboratory, Biomedical Engineering and Computational Science, Aalto University, Finland, [3]Cognitive Brain Research Unit, Institute of Behavioural Sciences, University of Helsinki, Finland
Time: 15:00

1. Background: While there is a substantial amount of studies on neural correlates of musical emotions, the dynamics thereof are less well understood. In particular, it has not been studied which areas are associated with continuously changing emotions during music listening. This question can be tackled using two main approaches: encoding, referring to prediction of neural activity from emotion ratings; or decoding, referring to prediction of emotion ratings from neural activity. Concurrent use of both approaches has been suggested to provide a means to corroborate the obtained results (Naselaris et al. 2011). Aims: We investigated neural correlates of perceived emotion during continuous listening to music. In particular, we aimed to locate areas involved in the processing of arousal and valence and assess the inter-subject consistency of the neural representations thereof.

2. Methods: The brain responses of participants were recorded using functional magnetic resonance imaging (fMRI) during continuous listening of three music pieces of different genres. Subsequently, the participants continuously rated the perceived emotional activity and valence of these stimuli. Following this, we used both encoding and decoding approaches to investigate the neural correlates of these emotion dimensions.

3. Results: The encoding analysis revealed perceived arousal to be positively associated with activation mainly in superior temporal gyri and negatively associated with activation in orbitofrontal and limbic areas. Valence, on the other hand, correlated positively with superior temporal gyri and primary motor cortex, and negatively with orbitofrontal and limbic areas. Subsequent decoding analyses revealed that perceived arousal could be predicted significantly better than valence, and that the right superior temporal gyrus was the core area for the prediction of arousal.

4. Conclusions: In terms of brain areas identified, the results are mostly in conformity with those previously obtained with more block-design-based studies. The differences in decoding accuracy between the two emotion dimensions suggest, in line with previous emotion research, a higher inter-subject consistency in the neural substrates for arousal than for valence.

CONTINUOUS SELF-REPORT ENGAGEMENT RESPONSES TO THE LIVE PERFORMANCE OF A POST–SERIALIST SOLO MARIMBA WORK

Mary Broughton[1], Emery Schubert[2], Catherine Stevens[3], Dominic Harvey[4]
[1]The University of Queensland, Australia, [2]University of New South Wales, Australia, [3]University of Western Sydney, Australia, [4]The University of Western Australia, Australia

1. Background: We investigate observers' continuous self–report engagement with a post–serialist solo marimba work performed live. We examine the relationship of intensity, time, and musical structure to observer engagement responses. Exposure to Western music develops familiarity, and implicit acquisition of knowledge about structural regularities and sequencing of materials. Where structural norms are absent, observers must acquire that knowledge during the performance.

2. Aims: This paper investigates relationships of intensity, musical repetition, and time with observers' continuous self-report ratings of engagement with an unfamiliar, post–serialist solo marimba work performed live. Four hypotheses are tested: i) engagement ratings and musical repetition scores relate positively, ii) engagement ratings and intensity values relate positively, iii) there is a relationship between engagement and time, iv) an interaction between musical repetition and time occurs — engagement ratings are higher with musical repetition occuring over time.

3. Method: Following a training session, 19 audience members continuously self–reported level of engagement on a bipolar, one–dimensional scale using the portable Audience Response Facility (pARF). Observer responses were sampled at a rate of 2Hz during the live performance of the Source, I by Toshi Ichiyanagi (duration four mins 45 sec). An audio recording was made at the same time. Each participant's responses and the physical intensity of the audio time series were smoothed with a five–second forward moving average. Musical repetition scores were deduced from musicological analysis according to an a priori coding system. A musical cell similar to a previous cell was scored higher than a cell containing new material.

4. Results: The mixed models approach used to analyse the data has been used to model change in longitudinal sociological data sets (Bickel, 2007). Although like regression, it distinguishes four structures: fixed effects, random effects, measurement error, and serial correlation. Intensity, musical repetition, time, and the musical repetition*time interaction significantly predicted engagement responses. The single predictors had positive, and the interaction a slight negative, relationship with engagement. Intercepts and slopes significantly varied across participants, and negatively and significantly covaried.

5. Conclusions: Engagement responses corresponded predominantly with intensity. Observers' engagement was enhanced with musical repetition, however time ameliorated the effect.

[7D] Perception of Pitch

LIÈGE 14:00- 16:00 Thursday 7 Aug 2014

A NEURODYNAMIC ACCOUNT OF RESIDUE PITCH

Karl Lerud[1], *Ji Chul Kim*[1], *Edward W. Large*[1]
[1]*University of Connecticut, Department of Psychology, USA*

Time: 14:00

Many types of auditory neurons are known to respond to the amplitude modulation of a carrier signal, and amplitude modulation is thought to be important in pitch perception. However, there are waveforms that have identical Hilbert envelopes while eliciting different pitches. One example is a complex of equally spaced higher harmonics of a missing fundamental shifted linearly up or down. The shift leaves the frequencies equally spaced by the same amount, but also inharmonic since they are no longer harmonics of a common fundamental. The perceived pitch of the shifted complex moves in the same direction as the shift, while leaving the amplitude envelope the same. The "pitch shift of the residue" shows that the fine structure of a waveform plays an important role in the auditory system's determination of its pitch, an essential feature of both music and language. Here we model the perception of residue pitch with networks of nonlinear neural oscillators. These networks are meant to mimic the most essential dynamics of the auditory system that give rise to the perception of pitch. We stimulated the auditory networks with pitch-shifted complex tones and compared the results to psychophysical studies. Dynamical analysis suggests that the perceived pitches of the stimuli are regions of resonance of the fine structure-locked portions of the responses. These results are explained as a consequence of highly nonlinear systems such as the auditory system.

DOES AUDITORY MASKING EXPLAIN THE HIGH VOICE SUPERIORITY?

Song Hui Chon[1], *David Huron*[1]
[1]*School of Music, Ohio State University, USA*

Time: 14:30

ABSTRACT

The majority of music in the world employs multiple concurrent parts. Among these parts, the upper-most part or voice often carries the melody. Considering the upward spread of auditory masking patterns, one might think the high-voice melodies are conflicting with masking theory. In this paper, we investigate the mutual masking effects of concurrent high- and low-pitched complex tones. In addition, we consider four types of spectral envelope patterns and discuss their influence on auditory masking.

1. INTRODUCTION

One of the most characteristic differences between music and speech is the typical number of concurrent sound sources involved. Speech usually involves social turn-taking, with a single speech stream alternating between the conversants. By contrast, although there is significant music-making involving a single stream (what musicians call "monophony"), the majority of the world's music-making involves multiple concurrent sound sources, and multiple concurrent auditory streams.

Among these concurrent parts, the melody, which is in general more important than the rest, is usually placed on the high voice or part. This high-voice melody practice is consistent with the fact that the changes are most easily detected in the highest stream (Zenatti, 1969; Palmer & Holleran, 1994; Crawley et al., 2002). This is known as the "High Voice Superiority." The high voice superiority has been examined quite extensively by Trainor and her team (Fujioka et al., 2005, 2008; Marie & Trainor, 2013) who reported that it might "result from the neurophysiological characteristics of the peripheral auditory system."

Auditory masking (ANSI, 1960) is defined as "1. The process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound," and "2. The amount by which the threshold of audibility for one sound is raised by the presence of another (masking) sound. The unit customarily used is the decibel." In other words, masking means that when two sounds are present at the same time, one sound may not be heard as well as it would have been in isolation, due to the existence of the other sound. There are two types of auditory masking, temporal masking and frequency masking, according to the domain of the proximity of two sounds. In this paper, we will only consider the frequency masking.

Frequency masking has been studied well in the context of auditory filter and critical bandwidth (Fletcher, 1940; Greenwood, 1961; Plomp & Levelt, 1965; Scharf 1970; Patterson 1976; Moore & Glasberg, 1983; Zwicker & Fastl, 1990). Critical band usually refers to a frequency range, within which two pure tones will interact and not resolve perfectly. The masking pattern, known as the spreading function or pattern, has been obtained by examining the degree of masking effect according to the placement of two pure tones along the length of a critical band. The masking pattern is not symmetric around the center of the critical band. Rather, it has a longer tail towards the higher frequency, which is known to be the "upward spread of masking."

This upward spread of masking means that the masking effect of a lower tone on a higher tone within the same critical bandwidth is greater than vice versa. The fact that a lower tone masks a higher tone more effectively seems to be contradictory with the High Voice Superiority, the fact that most melodies are on the high voice. Or do they really conflict with each other?

To answer this question, we implemented a computer simulation in MATLAB to examine pairs of complex tones and to determine which one masks the other better. We considered four different amplitude patterns to test their impact on masking effectiveness. Our hypothesis is that on average, higher-pitched harmonic complex tones will tend to mask the partials of comparable lower-pitched tones more than vice versa.

2. COMPUTATIONAL MODEL

A simulation, based on the psychoacoustic model in Bosi and Goldberg (2003), was implemented in order to compare the mutual masking between pairs of complex tones differing in pitch. The amount of masking will obviously depend on many factors, including the spectral envelope of the participating tones. Hence, we considered four types of spectral envelopes ("uniform," "increasing," "grand average" and "per-pitch average"; the last two are from Plazak et al., 2010). For each of the four envelopes, we examined their masking impacts on 81 pitches from B0 to G7. Stimuli were generated in MATLAB using sinusoids at the fundamental frequency plus 323 overtones for the first three envelope types. The sound files from Plazak et al. (2010) were used for the "per-pitch average" type.

In our simulation, we paired all possible non-unison tones spanning the range B0 to C7. That is, we began by generating complex tones for a pair of B0 with C1, each tone exhibiting the particular amplitude envelope pattern for each study. The complex tones were then used as input to a masking model. The masking model begins by carrying out a Fast Fourier Transform (FFT). The frequency and amplitude resolution of the Fourier analysis depends to some degree on whether the harmonics are integrally related to the FFT size. The FFT size we used was set to a resolution of 1 Hz, so in order to maintain good resolution we rounded the synthesized input frequencies so that they would be at one hertz integers. Hence, for example, a fundamental of 261.6 Hz was rounded to 262 Hz.

For each of paired complex tones, we calculated the masking pattern using the Terhardt spreading pattern (Terhardt, 1979). The masking envelopes for each of the lower and higher tones were individually computed. We then superimposed the masking envelope from the lower tone on the spectrum for the upper tone, and calculated the total residual power above the masked threshold. Similarly, we superimposed the masking envelope from the higher tone on the spectrum for the lower tone, and again calculated the total residual power above the masked threshold. Comparing the two total residual power values, the tone exhibiting the greater residual power was deemed to be lesser masked of the two tones. Hereafter we will refer to the lesser masked of the two tones as the "predominant" tone, and the greater masked tone as the "obscured" tone.

2.1. Study 1: Grand Average Pattern

In the first instance, we decided to limit ourselves to harmonic complex tones since these tones are most widely used in music. Musical tones exhibit highly variable spectral recipes. Even within a single musical instrument, the harmonic spectrum can differ significantly from one tone to the next and according to the method of sound generation.

Plazak, Huron and Williams (2010) calculated an average harmonic spectrum for 1,338 recorded musical instrument tones in the McGill University Master Samples (Opolko & Wapnick, 1987). The recordings included 23 common Western orchestral instruments, including piccolo, flute, oboe, English horn, clarinet, bassoon, contrabassoon, trumpet, trombone, tuba, French horn, violin, viola, cello, and contrabass. Sandell (1991) calculated harmonic spectra for each of these recorded tones. The amplitudes for each harmonic represent averages over the duration of the tone so time-variant fluctuations were ignored. The analyses were normalized so that the most energetic partial in each tone was defined as 0 dB. Plazak and colleagues calculated a grand-average harmonic spectrum for all 1,338 tones. We obtained the numerical amplitude data from Plazak for the synthesis of complex tones based on the grand average.

2.2. Study 2: Per-Pitch Average Pattern

One might reasonably object that, while the grand-average tone is representative of harmonic complex tones in general, in reality tones may exhibit patterns of spectral change with respect to register. The harmonic spectra of low-pitched instrumental tones differ somewhat from the comparable spectra for higher-pitched tones.

Accordingly, for Study 2, we made use of an "average instrument" as calculated by Plazak et al. (2010). In this case, Plazak and colleagues calculated an average harmonic spectrum for all instruments playing a given pitch. For example, the harmonic spectrum for C2 consists of the average spectrum for 12 sounds in the database. Similarly, the harmonic spectrum for C4 consists of the average spectrum for 34 sounds. In this way, Plazak et al. calculated the "per-pitch" average spectra for tones ranging from B0 to C7.

The synthesized versions of the average harmonic spectra from Plazak et al. (2010) are permanently archived in wav files at the Ohio State University's Knowledge Bank website (https://kb.osu.edu/), from where we downloaded the files. Since they were readily in the wav format, the files were read in MATLAB and the FFT results were inputted to the masking calculation model.

2.3. Study 3: Uniform Pattern

Both Studies 1 and 2 used approximations of real musical instrument tones. In both studies, the envelopes of all sounds exhibit a spectral roll-off in high frequency, which is found in many natural sounds. We then decided a priori to examine rather artificial cases, because we suspected that the spectral roll-off might contribute to the higher-pitched tone's masking of the lower-pitched tone. The first artificial pattern we considered is a uniform envelope, where all harmonics would have the same amplitude.

The stimuli were synthesized in the exactly same way as in Study 1, except that the amplitude was constant for all partials.

2.4. Study 4: Increasing Pattern

With the motivation of examining artificial envelope patterns and their impact on masking, we decided to examine another pattern that probably would not exist in nature: the increasing amplitude envelope. The amplitudes of the 324 partials were linearly increasing from 0.2 to 0.8, the values of which were arbitrarily determined a priori. In other words, for each complex tone, the amplitude of the fundamental frequency would always be 0.2 and the amplitude of the 323rd (highest) overtone would always be 0.8.

The stimuli were synthesized in the exactly same way as in Study 1, except that the amplitude was in a linearly increasing pattern.

3. RESULTS

The simulation results are summarized in Tables 1 and 2. Table 1 shows the number of occasions when the higher tone or the lower tone of a pair was predominant. The numbers in parentheses are the percentage of the total numbers of such occurrences in each category. In each of the four studies, there were a total of 3240 paired comparisons.

In Table 1, the Studies 1, 2 and 4 all have more cases of the higher predominant tone, which is consistent with our hypothesis. Study 3 with the uniform envelope pattern showed the opposite case with a more probability of a lower predominant tone. This "Low Voice Superiority" may

stem from the uniform amplitude for all partials, since all other three studies have an uneven (either increasing or decreasing) amplitude patterns in frequency spectra.

The numbers in Table 1 seem to suggest that for each study, there's a tendency of either the high voice superiority or the low voice superiority. However, the proportion of the high/low voice superiority (53% to 63%) appears to be not too different from 50%, which is a random decision based on equal probability. To examine that these results are statistically significant, meaning that they are different from a random determination based on a 50/50 chance, we calculated Chi-square values and the significant values (*p*-values), which are presented in Table 2. All four studies show significant Chi-square statistics, suggesting that the numbers in Table 1 were not generated by a random 50/50 decision.

Table 1: Summary of four studies in terms of the total number of occurrences of the higher tone masking the lower tone of each pair and vice versa. In parentheses are the percentages of the total numbers of occurrences.

Study	High predominant	Low predominant
Study 1: Grand Average	**1732 (53%)**	1508 (47%)
Study 2: Per-Pitch Average	**1793 (55%)**	1447 (45%)
Study 3: Uniform	1332 (41%)	**1907 (59%)**
Study 4: Increasing	**2030 (63%)**	1210 (37%)

Table 2: Chi-square test results of the four studies to compare the results in Table 1 with a random determination based on a 50/50 chance.

Study	χ^2	*p*
Study 1: Grand Average	7.7525	**.0054**
Study 2: Per-Pitch Average	18.5275	**<.00005**
Study 3: Uniform	51.6078	**< 10^(-12)**
Study 4: Increasing	105.4541	**< 10^(-12)**

4. DISCUSSION

In this paper, we simulated the masking effect of pairs of non-unison tones synthesized with four different envelope types. Using tones synthesized with more realistic recipes (Studies 1 and 2), the hypothesized high voice superiority was observed. Tones with increasing amplitude patterns (Study 4), which is one of the two artificial patterns we considered, also showed the high voice superiority. The other artificial case of uniform amplitude patterns (Study 3) generated tones that exhibited the low voice superiority, against our hypothesis. This opposite behavior observed in Study 3 from all the other studies might suggest that an uneven amplitude pattern may be the fundamental reason of the high voice superiority. We will consider in future other amplitude patterns to examine this conjecture.

Overall, the results can be interpreted that placing melodies on high voices do make sense in terms of the upward spread of auditory masking, especially considering more realistic envelope patterns used in Studies 1 and 2. Since we have considered only synthesized timbres so far, it might be worthwhile to include recorded samples of real instrument sounds and repeat the analysis.

So far, we did not inspect possible patterns in terms of pitch relations. Are there some effects of register, such that a higher tone within the same octave is a predominant masker, but if the pitch difference increases it becomes an obscured tone? This investigation may reveal some hidden masking patterns based on pitch and register, which composers may have implicitly known for ages.

This paper brings the high voice superiority to a new perspective. Trainor and the colleagues have conjectured that the high voice superiority may result from auditory periphery, although they could find the effect in the auditory cortex. Our results suggest that it might be a consequence of masking in cochlea of complex tones with uneven amplitude patterns. Cochlea is at an earlier and lower level than the auditory cortex in the auditory system. Hence, our findings might be what Trainor and the team has speculated. In summary, the high-voice melodies may have been, after all, a natural development based on the upward spread of auditory masking, which we are all born with.

5. REFERENCES

American Standards Association (1960). Acoustic Terminology SI, 1-1960. *American Standards Association, New York.*

Bosi, M., & Goldberg, R. E. (2003). Introduction to digital audio coding and standards. *Springer.*

Crawley, E.J., Acker-Mills, B.E., Pastore, R.E., & Weil, S., (2002). Change detection in multi-voice music: the role of musical structure, musical training, and task demands. J. Exp. Psychol. Hum. Percept. Perform. 28, *367-378.*

Fletcher, H. (1940). Auditory patterns. Reviews of modern physics, 12*(1), 47-55.*

Fujioka, T., Trainor, L.J., & Ross, B. (2008). Simultaneous pitches are encoded separately in auditory cortex: an MMNm study. Neuroreport 19, 361-366.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. J. Cogn. Neurosci. 17, *1578-1592.*

Greenwood, D. (1961). Auditory Masking and the Critical Band, J. Acoust. Soc. Am., 33, *484-501.*

Marie, C. & Trainor, L. (2013). Development of simultaneous pitch encoding: infants show a high voice superiority effect. Cereb. Cortex, 23, *660e669.*

Moore, B. C., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory filter bandwidths and excitation patterns. J. Acoust. Soc. Am., 74*(3), 750-753.*

Opolko, F. & Wapnick, J. (1987). McGill University Master Samples. (Compact discs OWCD 1-87, 2-87, 3-87). Montréal, Quebec: McGill University.

Palmer, C., & Holleran, S., 1994. Harmonic, melodic, and frequency height influences in the perception of multivoiced music. Percept. Psychophys. 56, *301-312.*

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. J. Acoust. Soc. Am., 59*(3), 640-654.*

Plazak, J., Huron, D., & Williams, B. (2010). Fixed Average Spectra of Orchestral Instrument Tones. Empirical Musicology Review, 5*(1), 10-17.*

Plomp, R. & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. J. Acoust. Soc. Am., 38, *548-560.*

Sandell, G. J. (1991). A library of orchestra instrument spectra. In Proceedings of the 1991 International Computer Music Conference. *San Francisco, CA: International Computer Music Association, 98-101.*
Scharf, B. (1970). Critical bands. In Foundations of modern auditory theory, *New York Academie, 157-202.*
Terhardt, E. (1979). Calculating virtual pitch. Hearing research, 1*(2), 155-182.*
Zenatti, A., 1969. Le développement génétique de la perception musicale. Monogr. Francaises Psychol. 17. *CNRS, Paris.*
Zwicker, E., & Fastl, H. (1990). Psychoacoustics: Facts and Models. *Springer-Verlag, Berlin Heidelberg.*

SPATIAL REPRESENTATIONS OF PITCH AMONG INDIVIDUALS WITH CONGENITAL AMUSIA

Xuejing Lu[1], Pin Zhang[2], Yanan Sun[3], William Forde Thompson[1]
[1]Department of Psychology, Macquarie University, Australia, [2]Medical Psychological Institute, Central South University, China, [3]Department of Cognitive Science, Macquarie University, Australia
Time: 15:00

ABSTRACT

Background: The link between pitch and space is not new. However, it remains unclear where the correspondence between pitch and spatial elevation comes from. One theory infers that people obtain this correspondence through statistical learning or language development. That is, pitch and space are represented separately but linked by association. Another theory postulates that pitch and space share common representational and processing resources. Evidence for either theory is lacking, suggesting a need for more research. The study of congenital amusia (CA), a pitch perception disorder, may shed light on this issue. If musical pitch and spatial processing draw on shared mechanisms, then pitch processing impairments in CA should be accompanied by an impaired capacity to represent pitch spatially. The present study examined the ability to represent pitch spatially in individuals with and without congenital amusia. **Method:** A speeded classification task was used to investigate the interaction between pitch height and spatial location. Participants with CA and matched controls were asked to indicate whether the tone they heard was high or low by pressing one of two extreme top and bottom keys as accurately and quickly as possible. In the congruent condition, the top key indicated that the pitch was high and the low key indicated that the pitch was low; the response options were reversed in the incongruent condition. Accuracy and reaction times (RTs) were analyzed. **Results:** For both groups, RTs were faster when the pitch height corresponded with that of the location of response key than when it did not (i.e., SRC effect), although the overall RT was longer for CAs than for controls. **Conclusions:** As predicted by associative representation theory, people with CA exhibit a preserved spatial representation of pitch, suggesting that this representation is independent of pitch discrimination.

1. BACKGROUND

The link between pitch and space is reflected in the linguistic metaphor of high or low pitches, and ascending or descending pitch movement to found many languages (Mudd, 1963; Patel, 2008). This analogy is not only reflected in language, but is explicit in Western notation, in which pitch is arranged spatially on a musical stave. The correspondence between pitch and space also can be found when perceiving pitch. An ascending tone sequence is usually perceived as ascending in space (Bregman & Steiger, 1980) and the estimation of pitch height is affected by the height of concurrent gestures (Connell, Cai, & Holler, 2012).

Although the connection between pitch and space is not well understood, it is well known that spectral information provides valuable information about the location of a sound source. For example, growls tend to be come from large creatures that navigate close to ground, whereas high-pitched sounds are most likely to be produced by small creatures that live in trees or fly. Thus, from an ethological perspective, it may be adaptive for organisms to form stable associations between pitch and spatial location.

The correspondence between pitch and horizontal space is typically only observed among musicians (Lidji, Kolinsky, Lochy, & Morais, 2007), but there are stable links between pitch and vertical space. One possibility is that people learn this correspondence through long-term exposure to regularities in the environment. A second possibility is that it is largely a linguistic convention. A third possibility is that representations of pitch and space are managed by common mechanisms such that pitch is inherently audiospatial in nature (Connell et al., 2012; Parkinson, Kohler, Sievers, & Wheatley, 2012; Walker et al., 2010).

The study of congenital amusia, a disorder that affects the processing of musical pitch, offers a unique strategy for examining this issue. To this end, the present study examined the correspondence between pitch and vertical space among people with congenital amusia. If musical pitch and spatial representation draw on shared mechanisms, then pitch-processing impairments in amusia may be accompanied by deficits in the capacity to represent pitch spatially.

2. METHOD

2.1. Paradigm

The correspondence between pitch height and spatial elevation can be readily observed in a speeded classification task, whereby pitches are classified as “high” or “low” by pressing on response options that are high or low in physical space. If a pitch-space correspondence exists, then performance on congruent trials (i.e., high pitches classified by selecting the high response option) should be better than performance on incongruent trials. Previous studies have confirmed that binary-choices are made faster when the stimulus (i.e. pitch height) corresponds with the vertical location of the response option than when it does not (Lidji et al., 2007; Rusconi, Kwan, Giordano, Umilta, & Butterworth, 2006). The phenomenon is known as stimulus-response compatibility (SRC), where dimensional overlap occurs between stimuli and responses.

2.2. Participants

People were diagnosed as amusic if they achieved a global score of 78% or lower on the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz, Champod, & Hyde, 2003). Eight (7F/1M) individuals with congenital amusia and 8 (7F/1M) controls participated. The pre-tests included Memory for Digits and Block Design subtests from WAIS-III, in order to control the working memory capacity and general spatial ability (see details in Table 1). All participants were right-handed and reported no history of auditory, neurological or psychiatric disorders. None had any formal musical education except regular music courses provided by schools. Written informed consent was obtained from all participants prior to the study. The experimental protocol was approved by the Ethics Committee of The Second Xiangya Hospital.

2.3. Procedure

Participants were tested individually in a quiet room. On each trial they were asked to put their two index fingers on the "6" and "SPACE" keys of the laptop's keyboard respectively in advance, and to fixate on the cross in the centre of the computer screen as a tone was presented via headphones. Participants were required to classify the tone they heard as "higher" or "lower" by pressing one of the response keys as accurately and quickly as possible. Only two pure tones were used in this task - a lower-pitched tone A4 (440Hz) and a higher-pitched tone A5 (880Hz). Each pure tone lasted 500ms and was generated by Audacity 2.0.3 with constant amplitude and 20ms onset and offset ramp. The deadline to respond was fixed at 2000 ms, including the duration of the tone itself. The inter-trial interval was 500ms. There were 2 blocks corresponding to the two response conditions (congruent and incongruent). In the congruent condition, the high-pitched tone was indicated by pressing the upper response key and the low-pitched tone was indicated with the lower response key. The incongruent condition had the reverse mapping. Each tone was presented 48 times in each block and trials were randomized with the stipulation that no more than four tones with same frequency ever occurred in a row. 8 practice trials were presented before each block. The order in which the blocks were presented to participants and hand position were counterbalanced among participants. Before commencing the task, participants were familiarized with the high- and low-pitched tones. Feedback was provided during the practice trials, but in the experimental trials feedback was only given for incorrect trials and for time-outs (no response within 2000 ms). Response times and accuracy were recorded.

Table 1: Participants' characteristics and standard scores in Memory for Digits and Block Design of WAIS-III, as well as the mean percentage of correct in MBEA and each subtest.

	Amusics (n=8)	Controls (n=8)	t-value (two-tailed)
Mean age (in year)	19.38	19.13	0.63
Mean educational year	13.13	13	1.00
Memory for Digits	9.88	13.38	1.84
Block Design	11.25	11.75	0.36
MBEA (%)			
Scale	65.00	86.67	5.13**
Contour	68. 75	88.33	4.05**
Interval	67.08	89.17	5.83**
Rhythm	69.58	85.42	2.53*
Meter	66.25	75.00	1.54
Memory	77.92	90.00	2.37*
Global score	69.10	85.76	6.26**

*p<0.05, **p<0.001

3. RESULTS

The mean reaction times (RTs) and error rates (ERs) from test blocks were entered into separate 2 x 2 repeated analyses of variance (ANOVA) respectively, with the within-subject variable of Congruence (congruent-incongruent) and the between-subject variable of Group (amusic-control). Responses were faster when low-pitched tones were indicated by the lower response-key and high-pitched tones were indicated by the higher response-key, as revealed by a marginally significant main effect of Congruence, $F=4.40$, $p=0.055$, $\eta^2=0.24$. Controls had faster RTs than amusics in both congruent and incongruent conditions at a group level, as revealed by the Group effect, $F=4.95$, $p<0.05$, $\eta^2=0.26$. Similarly, the main effect of Congruence on ER was significant, $F=13.64$, $p<0.05$, $\eta^2=0.49$, with lower ER for congruent trials. However, there was no group difference on ER (See Figure 1).



Figure 1: Mean RTs (left) and ERs (right) as a function of Congruence (congruent condition in blue; incongruent condition in red) in the amusic and the control group. Error bars represent +1 SEM.

4. DISCUSSION

The tendency to map auditory pitch onto visual elevation comprises one of the most robust cross-modal correspondences reported to date (Spence, 2011). Both perceived and imagined pitches can be translated into an analogical representation in the spatial domain (Elkin & Leuthold, 2011). The present study examined two alternative hypotheses on where the correspondence comes from by investigating pitch-space correspondence among people with congenital amusia. To our knowledge, this is the first study that observed the SRC effect in people with congenital amusia, suggesting that amusics preserve the ability to map pitch onto vertical space (also see Douglas & Bilkey, 2007). It supports the claim that pitch and space are represented separated but linked by association. Although impairment of musical pitch processing does not appear to affect the spatial representation of pitch, we also observed a group difference on overall reaction time, suggesting an overall reduction in the speed of judging pitch height. The longer reaction time may result from elevated thresholds of pitch discrimination (Foxton et al., 2004; Hyde & Peretz, 2004). Although the distance between the two tones in this study was large (12 semitones), amusics still may need more time to process pitch because of their deficits in fine-grained pitch perception. Alternatively, the two groups may both represent pitch spatially but in different ways (Spence, 2011). Regardless, our data indicate that individuals with congenital amusia have a preserved spatial representation of pitch, suggesting the representation of pitch is independent of low-level pitch processing, as predicted by associative representation theory.

5. REFERENCES

Bregman, A. S., & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of "what" and "where" decisions in audition. Perception & Psychophysics, 28, *539-546.*

Connell, L., Cai, Z. G., & Holler, J. (2012). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Paper presented at the Proceedings of the 34th Annual Conference of the Cognitive Science Society, Austin, TX.*

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. Nature neuroscience, 10, *915-921.*

Elkin, J., & Leuthold, H. (2011). The representation of pitch in auditory imagery: Evidence from S-R compatibility and distance effects. Journal of Cognitive Psychology, 23, *76-91.*

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. Brain, 127, *801-810.*

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. Psychological Science, 15, *356-360.*

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? Journal of Experimental Psychology: Human Perception and Performance, 33, *1189-1207.*

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone. Journal of Experimental Psychology, 66, *347-352.*

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. Perception, 41, *854-861.*

Patel, A. D. (2008). Music, language and the brain. *New York: Oxford University Press.*

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. Annals of the New York Academy of Sciences, 999, *58-75.*

Rusconi, E., Kwan, B., Giordano, B., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. Cognition, 99, *113-129.*

Spence, C. (2011). Crossmodal correspondences: a tutorial review. Attention, Perception & Psychophysics., 73, *971-995.*

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. Psychological Science, 21, *21-25.*

THE TRITONE PARADOX AMONG CHINESE CHILDREN AGED 12 AND 13

Diana Deutsch[1], Cong Jiang[2], Trevor Henthorn[1], Shibin Zhou[2]
[1]University of California, San Diego, USA, [2]Capital Normal University, Beijing, China
Time: 15:30

ABSTRACT

The tritone paradox is produced by a basic pattern that consists of an ordered pair of Shepard tones that are related by a half-octave (i.e., a tritone). When listeners judge whether such tone pairs ascend or descend in pitch, their judgments show systematic relationships to the positions of the tones along the pitch-class circle: Tones in one region of the circle are heard as higher and tones in the opposite region are heard as lower. However, listeners disagree as to which tones are heard as higher and which as lower; further, their perceptions vary with their languages or dialects, which in turn involve differing pitch ranges for speech. Studies on the tritone paradox have so far studied only non-tone language speakers, and teenage children have not so far been tested. This study explored perception of the tritone paradox in Chinese children aged 12 and 13 who were living in Beijing. All subjects spoke Mandarin, but the dialects that were spoken by their family members varied. Overall, a very strong tritone paradox was obtained. Pitch classes B and C tended to be judged as higher than the other pitch classes, with the plot of peak pitch classes falling off on either side of these. This finding probably reflects the most frequent pitch ranges of speech in Northern Chinese Mandarin dialects.

1. INTRODUCTION

The tritone paradox is a musical illusion that reflects an influence of speech on perception of musical patterns (Deutsch, 1991, 1992, 2007, 2013; Deutsch *et al.*, 1987, 1990, 2004, Ragozzine and Deutsch, 1994). The basic pattern that produces this illusion consists of an ordered pair of Shepard tones that are related by a half-octave (i.e., a tritone). So, for example, C might be presented followed by F#, or D followed by G#, and so on. Each tone consists of six sinusoids that stand in octave relation, and their amplitudes are determined by a bell-shaped spectral envelope (Figure 1). Since such tones lack the other harmonics that are needed to define a fundamental, their perceived heights are ambiguous, but their pitch classes are clearly defined.

The tritone paradox has two surprising features. First, when a given tone pair is presented (for example, D followed by G#) one listener might clearly hear an ascending pattern, while another listener might just as clearly hear a descending pattern instead. Yet when another tone pair is presented (for example, G followed by C#) the first listener now hears a descending pattern and the second listener hears an ascending one. As a second surprising feature, the way any one listener hears the tritone paradox generally varies in an orderly fashion with the positions of the tones along the pitch class circle: Tones in one region of the circle are heard as higher and those in the opposite region as lower.

To plot a listener's perception of the tritone paradox, tone pairs are presented in haphazard order, such that each of the 12 tones within the octave serves equally often as the first tone of a pair. The tones are produced under envelopes that are placed along four different positions along the spectrum, spaced at half-octave intervals. Averaging results over these different envelope positions enables control for overall effects of pitch height and also for the relative amplitudes of the components of the tones.

It has been found that the way the tritone paradox is perceived varies with the language or dialect to which the listener has been exposed. In addition, the dialects spoken by the listener's parents exert a particularly strong influence, indicating that the way the illusion is perceived is determined by a pitch class template that develops early in life (Deutsch, 1991, 2007, Deutsch *et al.*, 2004, Ragozzine and Deutsch, 1994). Deutsch (1991) compared two groups of subjects. The first group had grown up in California, and the second group in the south of England. The two groups differed statistically in their perceptions of the tritone paradox: Frequently when a subject from California heard a pattern as ascending a subject from the south of England heard the same pattern as descending; and vice versa. Further, Ragozzine and Deutsch (1994)

found that, among subjects who had grown up in the area of Youngstown, Ohio, the perceptions of those whose parents had also grown up in Youngstown differed substantially and significantly from those whose parents had grown up elsewhere in the United States. In another study, Deutsch *et al.* (2004) tested subjects whose first language was Vietnamese and who had moved to California. The first, older, group had arrived in the U.S. as adults and the second, younger, group had arrived in the U.S. as infants or children. The perceptions of these two groups of subjects were very similar to each other, and both differed substantially from those of a third group, who were native speakers of Californian English and whose parents were also native speakers of Californian English. In yet another study, Deutsch (2007) found a significant correlation between the way children and their mothers heard the tritone paradox. This was true even though the mothers had grown up in different geographical regions, and so perceived the tritone paradox differently from each other.

The present study explored perception of the tritone paradox in a group of children aged 12 and 13 who were living in Beijing, China. While the majority of the subjects had spent all their lives in Beijing, a substantial minority had lived elsewhere, in a variety of provinces, and their mothers and fathers had lived in many different regions of China. We were interested in two basic issues: First we sought to determine how strongly these children perceived the tritone paradox. Second, we sought to determine whether, taking the group as a whole, the tritone paradox still showed an effect of how the pitch class circle was perceptually oriented with respect to height, taking the group as a whole.

2. METHOD

2.1. Stimulus patterns.

The stimulus patterns were taken from Deutsch's CD *Musical illusions and paradoxes* (1995). All tones consisted of six sinusoids that stood in octave relation, and the amplitudes of the sinusoids were determined by a fixed, bell-shaped spectral envelope. Figure 1 displays, as an example, a spectral plot of the tone D-G# generated under an envelope that was centered on C_5. In order to control for possible effects of the relative amplitudes of the sinusoids, the tones were generated under envelopes that were placed at four different positions along the spectrum, spaced at ½ octave intervals. Specifically, the envelopes were centered at 262 Hz (C_4), 370 Hz ($F\#_4$), 523 Hz (C_5) and 740 Hz ($F\#_5$). Twelve tone pairs were produced under each of the four spectral envelopes; these consisted of the pitch-class pairings C-F#, C#-G, D-G#, D#-A, E-A#, F-B, F#-C, G-C#, G#-D, A-D#, A#-E, and B-F. Four different orderings of these 12 pitch class pairings were created. These orderings were random, except that no consecutive pairings contained tones of the same pitch class. (For example, C-F# followed by F#-C was not allowed.) The tones were presented in blocks of 12. Each block consisted of tones that were generated under one of the four spectral envelopes and contained one example of each of the 12 pitch-class pairings – so there were 16 blocks altogether. All tones were 500 ms in duration, and the tone pairs were separated by 5-sec inter-trial intervals. The blocks were separated by 1-min pauses, except that there was a 5-min break between the eighth and ninth blocks. The subjects listened to the sound patterns from the CD through loudspeakers at a comfortable listening level.

2.2. Subjects

Potential subjects were played 48 examples of tritone pairs produced by pure tones, and were asked to determine for each pair whether it ascended or descended in pitch. Forty-seven subjects were selected on the basis of obtaining a score of at least 90% on this test, and without regard for musical training. No subject reported any hearing problems. The subjects were 20 male and 27 female students at Xinghua Middle School in the Daxing district of Beijing. All subjects spoke Mandarin. Their average age was 12.32 years (range 12-13 years). Thirty-seven subjects reported having had private instrumental or vocal music lessons. For these subjects, the number of years of musical training averaged 4.5 years (range 1 – 10) years. Thirty-two subjects had been born and lived all their lives in Beijing; however, the subjects' parents and grandparents were from diverse provinces within China, as were their grandparents (who frequently served as caregivers).

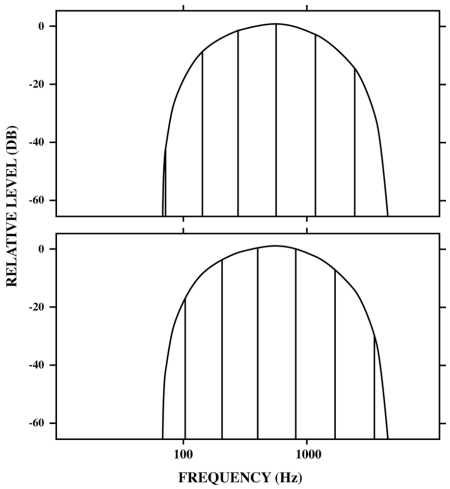


Figure 1. Spectral plot of a tone pair that produces the tritone paradox. Here the spectral envelope is centered at C_5. Upper graph represents a tone of pitch class D; lower graph represents a tone of pitch class G#.

2.3. Procedure

The subjects were tested during class time in a quiet room. They were first administered a questionnaire that enquired into their musical background, what languages they spoke, where they had lived, and where their parents and grandparents had lived. Following this, they were played 48 examples of tritone pairs consisting of pure tones, and were asked to determine for each pair whether it ascended or descended in pitch. It was determined that only the data from students who obtained a score of at least 90% correct on this pure tone test would be analyzed in the

tritone paradox experiment, which followed. In the experiment, six tritone pairs were first presented for practice without feedback, and these were followed by the 16 experimental blocks. On each trial one of the tone pairs was presented, and the subject indicated in writing whether it formed an ascending or a descending pattern.

3. RESULTS

For each subject, the percentage of judgments that a tone pair formed a descending pattern was plotted as a function of the first tone of the pair. In addition, and the percentage of judgments that a tone pair formed an ascending pattern was plotted as a function of the second tone of the pair. These two sets of judgments were averaged, producing a plot of the percentage of times that each pitch class was heard as the higher of a pair. Figure 2 shows the plots produced by three of the subjects. It can be seen that all subjects showed strong relationships between pitch class and perceived height; however the form of this relationship varied clearly across subjects.

To estimate the prevalence of tritone paradox in the subject population as a whole, the following procedure was employed. First, it was determined for the scores of each subject separately whether the pitch class circle could be bisected such that none of the scores in the upper half of the circle was lower than any of the scores in the lower half. The data of 39 of the 47 subjects met this criterion. Next, a baseline estimate was obtained of the probability of obtaining such a result by chance: The scores of each subject were subject to 1 million permutations, and the probability of meeting this criterion was thereby evaluated. Averaged across subjects, the probability of a subject meeting this criterion by chance was 5.29%. The combined probability that 39 of 47 subjects would meet this criterion is therefore vanishingly small, so we conclude that the tritone paradox exists to a highly significant extent in this population.

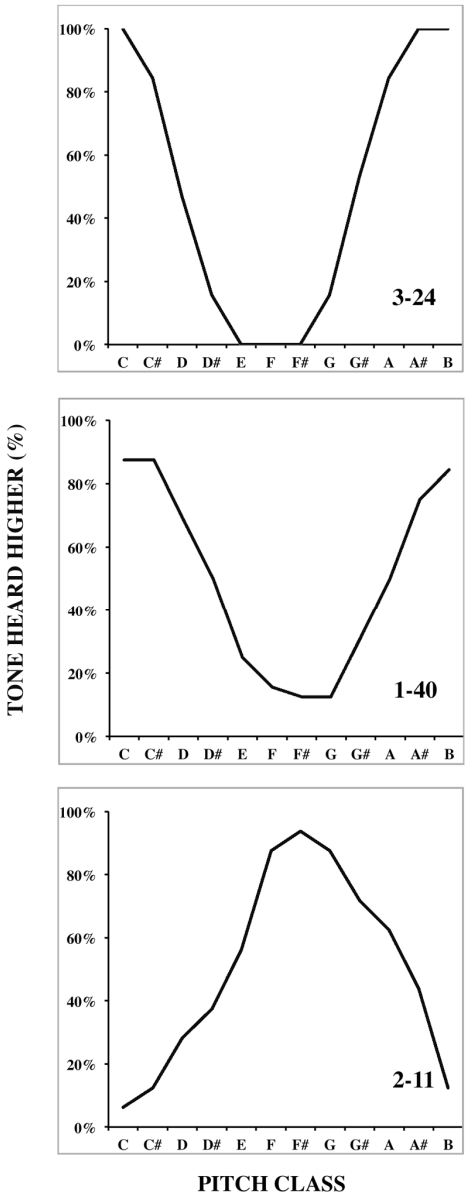


Figure 2. Percentages of trials in which a tone was heard as the higher of a pair, plotted as a function of the pitch class of the tone. The data from three individual subjects are displayed.

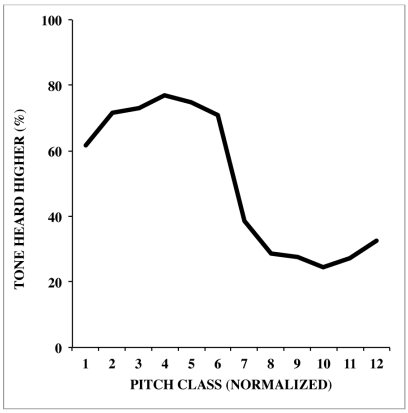


Figure 3. Percentages of trials that a tone was heard as the higher of a pair, plotted as a function of the pitch class of the tone. The data were averaged across all subjects and across all four spectral envelopes, and the orientation of the pitch class circle was normalized across subjects.

We next explored the form of relationship between pitch class and perceived height in this population as a whole. The orientation of the pitch class circle was normalized across subjects using the following procedure. First, for each subject the pitch class circle was bisected so as to maximize the difference between the averaged scores within the upper and lower halves. The circle was then oriented so that the line of bisection was horizontal. The data were then retabulated, with the leftmost pitch class of the higher half of the circle taking the first position, its clockwise neighbor taking the second position, and so on. The normalized data were then averaged across subjects. The resultant graph is shown in Figure 3, and it can be seen that the overall relationship between pitch class and perceived height for this group of subjects was very orderly.

Finally we determined whether, within this population, certain pitch classes were perceived as higher, and others as lower. It had been found, for example, that native English-speaking Californians tended to hear pitch classes from B to D# (moving clockwise along the pitch class circle) as higher, and those from E to A# as lower. In contrast, subjects from the South of England tended to hear pitch classes F#, G, and G# as higher and the other pitch classes as lower (Deutsch, 1991). To establish the form of relationship between pitch class and perceived height for the present group, for each subject the pitch classes that stood at the top of his or her pitch class circle (i.e., at positions 3 and 4 on the normalized plot) was determined. These were designated peak pitch classes.

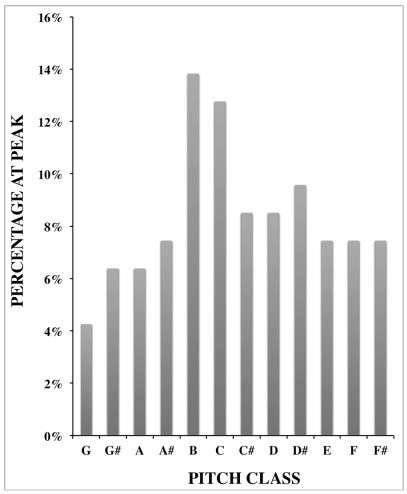


Figure 4. Distribution of peak pitch classes within this subject population, consisting of children aged 12 – 13 living in Beijing.

Figure 4 shows the distribution of the peak pitch classes in the subject population. It can be seen that B and C occurred more often as peak pitch classes than did the others. However, the distribution is nevertheless quite widespread, and this probably reflects the strongly heterogeneous linguistic background of the present group of subjects.

4. DISCUSSION

This paper provides the first study of perception of the tritone paradox in a large group of school-aged children, and also in speakers of tone language – specifically Mandarin. We found that the illusion is highly prevalent in the present population – at least as prevalent as found earlier by Deutsch *et al.* (1987) in a group of Californian English-speaking undergraduates. This high prevalence – particularly considering the young age of the subjects, may well be due to the importance of pitch height in the enunciation of Mandarin tones.

In the present experiment, it was found that pitch classes in the range from B to D (moving clockwise) were perceived as higher than the other pitch classes. In earlier work (Deutsch *et al.*, 1990, 2004) it has been hypothesized that the template that determines perception of the tritone paradox also determines the pitch range of the listener's speaking voice. And indeed, informal observations by one of us (DD) have indicated

that Tone 1 (high-flat) is likely to be pronounced by people from Beijing in the range from B to D (moving clockwise). This hypothesis, linking perception of the tritone paradox to the enunciation of Mandarin tones, will be explored formally in further research.

The present findings – in addition to those of earlier studies on the tritone paradox – have implications for theories of absolute pitch. Since judgments of the tritone paradox are strongly influenced by pitch class, subjects who obtain this illusion must be using a form of absolute pitch in making their judgments. In earlier studies of this illusion, non-tone language speakers were employed, and very few if any of these subjects would have possessed absolute pitch. For speakers of non-tone language, then, the tritone paradox shows that absolute pitch must be present in implicit form – as has been argued by Deutsch and colleagues in earlier papers (see, for example, Deutsch *et al.*, 1987; Deutsch, 2013). The present subjects were not tested for absolute pitch, and it is known that this faculty is highly prevalent among Mandarin speakers with musical training (Deutsch, 2013). However, the high prevalence of the tritone paradox in the present study makes it very improbable that explicit absolute pitch (in the form of verbal labeling) was here involved to a significant extent.

5. REFERENCES

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. Music Perception, 8, *335-347*

Deutsch, D. (1992). Paradoxes of musical pitch. Scientific American, 267, *88-95*

Deutsch, D. (1995) Musical Illusions and Paradoxes, *[CD with booklet] La Jolla: Philomel Records.*

Deutsch, D. (2007). Mothers and their offspring perceive the tritone paradox in closely similar ways. Archives of Acoustics, 32, *3-14*

Deutsch, D. (2013). Absolute pitch. In D. Deutsch (Ed.) The psychology of music, 3rd Edition, *141-182, San Diego: Elsevier*

Deutsch, D., Henthorn T. & Dolson, M. (2004). Speech patterns heard early in life influence later perception of the tritone paradox. Music Perception, 21, *357-372*

Deutsch, D., Kuyper, W. L. & Fisher, Y. (1987). The tritone paradox: Its presence and form of distribution in a general population. Music Perception, 5, *79-92*

Deutsch, D., North, T. & Ray, L. (1990). The tritone paradox: Correlate with the listener's vocal range for speech. Music Perception, 7, *371-384*

Ragozzine, F. & Deutsch, D. (1994). A regional difference in perception of the tritone paradox within the United States. Music Perception, 12, *213-225*

[7E] Schemata

LOS ANGELES 14:00- 16:00 Thursday 7 Aug 2014

DOES HARMONY AFFECT SCALE DEGREE QUALIA?: A CORPUS STUDY INVESTIGATING THE RELATIONSHIP OF SCALE DEGREE AND HARMONIC SUPPORT

Claire Arthur[1]
[1]*Ohio State University, USA*
Time: 14:00

ABSTRACT

The topic of expectancy in music has long been a topic of scholarly interest. Studies of musical expectancy bear important implications for theories of perception, and can help further our understanding of learning processes. A project is reported which attempts to investigate the interrelationship of melodic and harmonic expectancy via a corpus analysis. Findings from the corpus analysis related to harmonic influences upon melodic tendencies lay a foundation for a perceptual experiment that probes the effects of harmony on scale-degree *qualia*.

1. INTRODUCTION

What is the effect of harmony on our perception of scale-degree? When training students to hear scale-degree, teachers often rely on the use of qualia. For instance, scale-degree 7 ('ti') might be described as 'feeling like it wants to move upwards'. But where do scale-degree qualia come from? Huron (2006) discovered that actual scale-degree tendencies from melodies (specifically, from the Essen folksong collection) closely matched our intuitions about melodic expectancies, and thus argued that scale-degree qualia arise from statistical learning. But do scale-degree qualia arise solely from exposure to melodic patterns, as Huron posits, or does harmony play a role? Understanding the influence of harmony on the perception of scale-degree is an important pursuit, since the perceptual effects may have significant implications for music pedagogy and the study of musical expectations.

2. BACKGROUND

Although any scale-degree can function in multiple harmonic contexts, certain scale-degrees are affiliated more closely with one harmony than another. For instance, scale-degree 7 ("ti") is associated more with dominant harmony (V) than with the mediant (iii). Since dominant typically moves to tonic, scale-degree 7 will have a relatively strong schema ('ti goes to do'). Not all scale-degrees and harmonies have the same need for resolution, however. Furthermore, the need for melodic resolution can vary with harmonic context. Scale-degree 4 (fa), for instance, might behave differently depending on whether it is supported by V7 or by IV. The common melodic expectancy is for scale-degree 4 (fa) to fall to scale-degree 3 (mi). Does this melodic expectancy arise simply because within melodies "fa" tends to preceed "mi" regardless of harmonic context? Or is this melodic expectancy tied to the voice-leading and resolutional expectancies inherent in the dominant 7th (V7) harmony?

Thus, the question of interest is: do melodic expectations arise solely from statistical inference of melodic behavior, independent of harmonic context? Or does melodic behavior differ depending on the underlying harmonic support? It is hypothesized that harmony (in particular, harmonic function) interacts with melodic expectations, and that this would have an effect on scale-degree qualia.

It is proposed that the phenomenon of scale-degree qualia is strongly tied to melodic tendency, which is in turn influenced by three primary factors: the strength of a scale-degree's association to a particular harmony, the predictability of scale-degree behavior within a particular harmonic context, and the proportion of associated harmonies that fall under the same harmonic function.

3. METHODS

3.1. Sampling

A corpus of classical music is analyzed to determine the probabilities of harmonic support for each scale-degree (e.g. likelihood of scale-degree 7's association with vii, V, and iii). The pieces used in the corpus were inspired by a subset of pieces from the Norton Anthology for Musical Analysis (5th edition), and includes a variety of styles and composers. Pieces were first chosen based on their availability in computer-readable format, and were prioritized based on an optimum level of distinction (or 'separability') between melody and harmonic accompaniment. In order to ensure some degree of consistency of harmony and voice-leading practices among the pieces to be sampled, only those pieces composed between the Baroque and Romantic period were used (roughly 1600-1850). The corpus was prepared using the Humdrum Toolkit, with some of the harmonic analyses already encoded and available through the kern scores database. The remaining harmonic analyses were added manually. In order to have enough data to be able to make generalizable conclusions, certain decisions had to be made that would simplify the harmonic context. For instance, there may be a difference in melodic behavior for scale-degree depending on whether the supporting harmony is in root position or not. However, classifying all harmonies separately based on both Roman numeral *and* inversion would divide the data too much, and make it more difficult to discover patterns. Thus, in this analysis all harmonies were represented with a simple roman numeral without any inversion data. Other simplifications included 'collapsing' seventh chords in with triads, and ignoring durational values of both harmonies and melodic tones.

3.2. Analysis

The analysis itself first consisted of tallying the proportions of harmonic support for each scale-degree in isolation. Then, the natural distribution of harmonies were taken into account. That is, since the natural usage of harmony is not random, with some harmonies used more than others, it is possible, for instance, that a particular scale-degree will rarely be paired with a given harmony. This could be a result of the rarity of a harmony in general (i.e. ♭II), or it could be a result of an 'avoidance' of a particular harmonic setting. Next, the first-order probabilities were calculated independently for scale-degree, and for harmony. That is, given some scale-degree, x, which scale-degree is likely to follow it? Finally, these two-note scale-degree 'windows' were compared against their harmonic support to investigate whether the harmonic support appears to influence the melodic behavior.

Results from the corpus analysis will be used to inform a model of scale-degree along a continuum of 'strong/weak' association with a particular harmony/harmonic progression. This model will then be tested in a future perception experiment.

There are perceptual and pedagogical implications for both melodic expectation and for scale-degree qualia that can be tested in an experimental setting. First, scale-degrees that are more plurivalent with regards to harmonic setting may have weaker schemata, which suggests that in a scale-degree identification tasks, a scale-degree that is not strongly associated with any particular harmony could more likely be mistaken for some other 'weakly' associated scale-degree. Second, if the predictability of melodic tendency is (harmonic) context-dependent, one would expect that in a judgment/prediction task listeners would infer different melodic trajectories depending on harmonic context.

4. REFERENCES

Arden, B. (2006.) How the timing between notes can impact musical meaning. Bologna, Italy: International Conference on Music Perception and Cognition.

Bharucha, J., & Krumhansl, C. (1983). The representation of harmonic structure in music: Hierarchies of stability as a of context. Cognition, *13, 63-102.*

Boltz, M. (1989). Perceiving the end: Effects of tonal relationships on melodic completion. Journal of Experimental Psychology: Human Perception and Performance, *15, 749-761.*

Butler, D., & Brown, H. (1981). Diatonic trichords as minimal cue cells. In Theory Only, *5/6, 39-55.*

Butler, D., & Brown, H. (1984). Tonal structure versus function: Studies of the recognition of harmonic motion. Music Perception, *2, 6-24.*

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. Psychomusicology: A Journal of Research in Music Cognition, *1/1, 12-29.*

Cross, I. (1998). Music analysis and Music Perception. Music Analysis, *Vol. 17, No.1, 3-20.*

Cuddy, L. L. & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. Perception & Psychophysics, *57/4, 451-462*

De Clercq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. Popular Music, *Vol. 30, No. 1, 47-70.*

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. Psychological Review, *88/6, 503-522.*

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. Psychological Review, *85/4, 341-354.*

Hansberry, B. (2013). What are scale-degree qualia?: An assessment of cognitive science and a philosophical account. Unpublished manuscript, Columbia University, NY.

Hollerman, S., Jones, M. R., & Butler, D. (1995). Perceiving implied harmony: The influence of melodic and harmonic context. Journal of Experimental Psychology, *21, 737-753.*

Huron, D. (2006). Are scale degree qualia a consequence of statistical learning? Bologna, Italy: International Conference on Music Perception and Cognition.

Huron, D. (2006). Sweet Anticipation. *Cambridge, Massachusetts: MIT Press.*

Krumhansl, C. (1991). Music psychology: Tonal structures in perception and memory. Annual Review of Psychology, *42, 277-303.*

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. Cognitive Psychology, *11/3, 346-374.*

Krumhansl, C. L., & Kessler E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation in music. Psychological Review, *89/4, 334-368.*

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. Journal of Experimental Psychology: Human Perception and Performance, *5, 579-594.*

Margulis, E. H. (2005). A model of melodic expectation. Music Perception, *22, 663-714.*

Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. Perception & Psychophysics, *70/5, 841-852.*

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. Journal of Mathematics and Music: Mathermatical Approaches to Music Theory, Analysis, Composition and Performance, *5/1, 35-53.*
Rohrmeier, M., & Cross, I. (2008). Statistical properties of tonal harmony in Bach?s chorales. Sapporo, Japan: International Conference of Music Perception and Cognition.
Schellenberg, E. G. (1997). Simplifying the implication -realization model of melodic expectancy. Music Perception, *14, 295-318.*
Schellenberg, E. G., Bigand, E., Poulin-Charronnat, B., Garnier, C. & Stevens, C. (2005). Children's implicit knowledge of harmony in Western music. Developmental Science, *Vol.8, No. 6, 551-566.*
Schmuckler, M. A. (1989). Expectation in music: Investigation of melodic and harmonic processes. Music Perception, *7, 109-150.*
Schmuckler, M. A., & Boltz, M. G. (1994). Harmonic and rhythmic influences on musical expectancy. Perception & Psychophysics, *53/3, 313-325.*
Temperley, D. (2004). The basic cognition of musical structures. *Cambridge, Massachusetts: MIT Press.*
Temperley, D. (2008). A probabilistic model of melody perception. Cognitive Science, *32, 418-444.*
Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. Psychological Review, *107/4, 885-913.*
Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy. Music Perception, *16, 99-117.*
Tymoczko, D. (2011). A geometry of music: harmony and counterpoint in the extended common practice. *New York: Oxford University Press.*
Unyk, A. M., & Carlsen, J. C. (1987). The influence of expectancy on melodic perception. Psychomusicology: Music, Mind & Brain, *Vol. 7, No. 1/3, 3-23.*

HUMANE TEMPO VS. METRONOMIC TEMPO

So-Young Lee[1], Elizabeth Park[1]
[1]Elizabeth Park Center for Professional Studies of Music Performance, Korea, Republic of
Time: 14:30

ABSTRACT

What would Bach say about my tempo of his Prelude?

How does a player decide upon a particular tempo, and to keep or not to keep that tempo throughout the entire piece? With the invention of metronome, tempo directives are even given with specific metronome numerical markings in the score. Even so, performers tend to sway away from metronomic tempo. We might think that it is natural phenomenon since the player is not a machine that can function precisely with the mechanical regularity. We also need to pay attention to the fact that it is hard to ignore performers' interpretation of tempo. Thus, can we say players intentionally choose a particular tempo based on subjective reading of the score to keep their musical character which, they believe, might be tarnished when they just follow metronome? Or can we say that players simply fail to follow written metronomic tempo and, thus, should work until they can play precisely along with metronome?

In an effort to answer these questions, we have audience-surveys for their preferences between the two different ways of performing: one with metronomic tempo and the other with humane tempo (performers' educated choice of tempo which is more in tune with the performers' natural expressive desire). Audience is composed of three separate groups; 1.Professional Musician 2.Music Connoisseurs 3.General Public excluding the other two. Performers play excerpts from each different period (Baroque, Classic, Romantic, Modern) with metronomic tempo and humane tempo. Audiences mark their preferences without knowing which is played with metronomic tempo or humane tempo.

Whether the three audience groups have similar preferences or not, the survey result will be of great interest and value to the performers for enhancing the quality of their performances and maximizing the audience satisfaction.

1. INTRODUCTION

Classical Music Concert Attendance has been decreasing.24 There may be numerous reasons for this trend. If one of the reasons is the audience find the classical music concert unappealing and boring, what would be the possible cause? Classical music, after all, must have its charms, given the fact that it has been played continuously for several centuries. However, people today are not living with the same level of stimulation as their predecessors did a hundred years ago, having many more attractions and forms of entertainment. Based on the assumption that classical music is gradually losing popularity among modern audiences who are accustomed to a more stimulating environment, we came up with a hypothesis that it may have to do with *tempo*, as one of the many possible causes of this phenomenon.

Since the advent of the metronome, composers have inserted tempo markings into scores as their channel of communication with performers. Out of respect to the composers, performers seek not to miss even the slightest hint from the score. As for the tempo, performers feel they ought to follow the suggested tempo markings as something absolute and employ a metronome. They work hard to keep in time with the metronome constantly and try not to deviate. Nonetheless, more often than not, the chances are high that the audience would still feel bored.

This brings us to the basic question: What does the audience feel about different tempos? In an effort to provide objective data in this respect, a survey was conducted on the audience's preferences between music played in a *metronomic* tempo and in a *humane* tempo which reflects the emotional reactions of the performer to the performed work at that given moment. We believe, by incorporating the survey results (i.e. data) into how they play, performers will be able to improve the quality of their practices and performances.

2. PREFERENCE SURVEY ON TEMPOS AND ITS RESULTS

To examine the audience's tempo preferences, performers were asked to play with a metronome and with a humane tempo25 (i.e. selectively deviating from the metronomic tempo) so the audience could listen to both and choose their preference. The audience was divided into three

24 A New Survey Finds a Drop in Arts Attendance –NY Times Sept 26, 2013

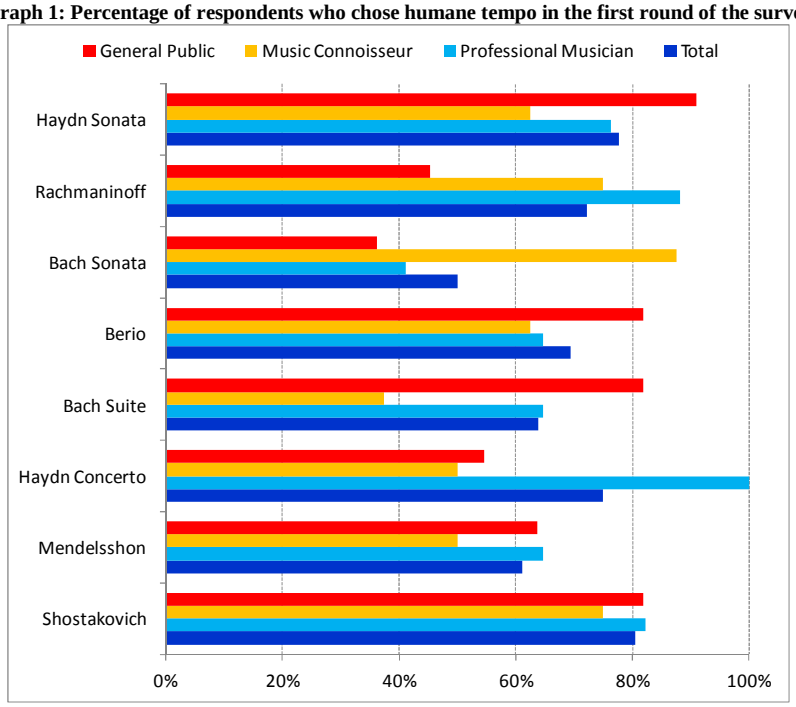
25 "Humane tempo" in this paper refers to a tempo that feels humane. It respects the composer's tempo markings yet reflects the emotional reactions of the performer to the performed work at that given moment.

groups of general public, music connoisseurs, and professional musicians. Performers played pieces from four different eras—Baroque, Classical, Romantic, and Modern—for two and a half minutes each in metronomic and in humane tempos to find out what each audience group preferred.

The survey was carried out in two rounds: In the first round, a performance seminar was held—with three groups of performers (piano, bass, and cello) playing a total of eight pieces live—to determine the audience's preferences. In the second round, a cellist (i.e. the author) and a pianist recorded the same repertoire as the first round, which three audience groups listened to on a CD player to decide which they liked the best.

In the live performances, elements of the live performance —encompassing not just what they listened to but also what they saw, background information of the pieces, and other complex elements of live performance—were reflected in the survey results. When it came to the recordings, on the other hand, they heard nothing but the sounds from the speakers, with all the other factors set aside. The situational differences affected the survey results, which are represented in the graph and chart below.

Graph 1: Percentage of respondents who chose humane tempo in the first round of the survey



Piece	11 General Public		8 Connoisseurs		17 professional musicians		36 Respondents in Total	
	M	H	M	H	M	H	M	H
Haydn Piano Sonata	9%	91%	38%	62%	24%	76%	22%	78%
Rachmaninoff Piano Variation	54%	46%	25%	75.%	12%	88%	28%	72%
Bach Viola da Gamba Sonata	64%	36%	13%	87%	59%	41%	50%	50%
Berio Bass Solo	18%	82%	38%	62%	35%	65%	31%	69%
Bach Cello Suite No.2	18%	82%	62%	38%	35%	65%	36%	64%
Haydn Cello Concerto	45%	55%	50%	50%	0%	100%	25%	75%
Mendelssohn Cello Sonata	36%	64%	50%	50%	35%	65%	39%	61%
Shostakovich Cello Sonata	18%	82%	25%	75%	18%	82%	19%	81%
Total	**33%**	**67%**	**37%**	**63%**	**27%**	**73%**	**31%**	**69%**

Chart 1: Percentage of respondents who chose metronomic and humane tempos in the first round of the survey

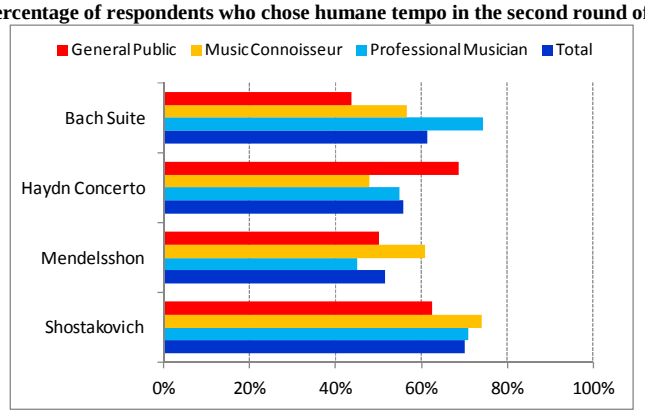
In the first round of the survey, three groups of performers played a total of eight pieces from different eras. The 36-member audience (comprised of 11 general public, 8 music connoisseurs, and 17 professional musicians) listened to the music played in metronomic and in humane tempos, without being told which tempo was used. The performers also tried to stay as neutral as possible without reflecting a personal preference of a certain tempo. The results varied slightly by the pieces: 31% of the respondents preferred music played with a metronomic tempo, 69% with humane tempo.

Interesting issues were found in the survey results: As for Bach's Viola da Gamba Sonata No. 2, 3rd mov. played by a double bassist, the difference was stark, as 87% of the connoisseurs chose the humane tempo, while only 41% of the professional musicians chose the same. The overall outcome across the three groups, however, was fairly even at 50/50. Performers need to pay close attention to the different tempo preferences between the connoisseurs and professional musicians.

In the case of Haydn Cello Concerto No. 1, 1st mov. played by a cellist, the preference among connoisseurs was even at 50/50. Among general public, 55% chose the humane tempo, while it was chosen by 100% of professional musician. As the cellist who played the piece live, the

author made a small mistake while playing with the metronome, which seemed to have no significant impact on the preferences of the general public and connoisseurs. The results suggest, however, that a performer's mistakes may affect significantly when professional musicians determine their preferences.

Graph 2: Percentage of respondents who chose humane tempo in the second round of the survey



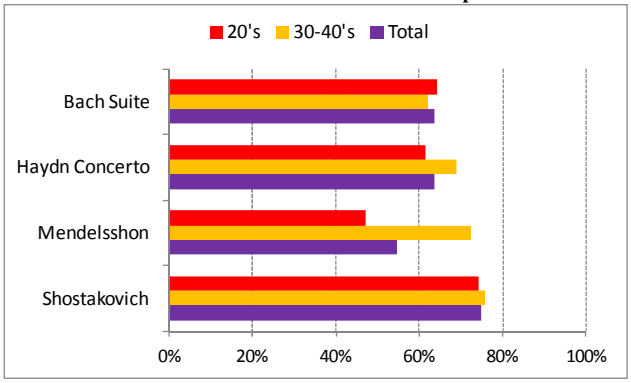
The results from the second round of the survey also varied across the three groups (16 general public, 23 music connoisseurs, and 31 professional musicians). A total of 70 respondents listened to four pieces and expressed their preferences. Of them, 40% chose the metronomic tempo and 60% the humane tempo, meaning the gap between the two tempos decreased remarkably compared to 31% vs. 69% in the first round of the survey. In the live performance, the audience seems to have found the humane tempo more appealing as it better reflects the musical intention of the performers, the body movements made during the performance, reverberations in the hall, and other elements that cannot be reproduced through recordings. Although the same performers have played the same pieces for both rounds, their performances could not have been exactly the same. The gap in preferences between the live performances and recordings is great. The gap between the metronomic and humane tempos narrowed in the recording-based survey, as the humane tempo appealed more to the audience in the live performances and the metronomic tempo was more preferred when it came from a machine (i.e. CD player).

Piece	16 General Public		23 Connoisseurs		31 Professional Musicians		70 Respondents in Total	
	M	H	M	H	M	H	M	H
Bach Cello Suite No.2	56%	44%	44%	56%	26%	74%	39%	61%
Haydn Cello Concerto	31%	69%	52%	48%	45%	55%	44%	56%
Mendelssohn Cello Sonata	50%	50%	39%	61%	55%	45%	49%	51%
Shostakovich Cello Sonata	38%	62%	26%	74%	29%	71%	30%	70%
Total	**44%**	**56%**	**40%**	**60%**	**39%**	**61%**	**40%**	**60%**

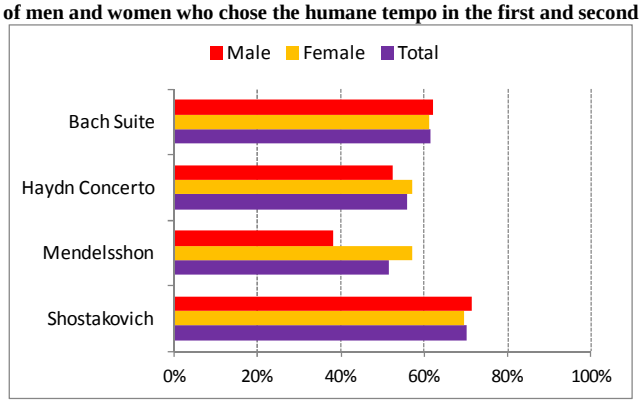
Chart 2: Percentage of respondents who chose metronomic and humane tempos in the second round of the survey

The results from the preference survey were reclassified by age and gender, and the outcome suggests the similar levels of preferences for the humane tempo. In the case of Mendelssohn's Cello Sonata No. 2, 3rd mov. Adagio, however, only 47% of those in their 20s chose the humane tempo, compared to 73% of those in their 30-40s. The percentage also differed by gender: 38% for men vs. 57% for women. This shows that younger listeners in their 20s and men prefer the metronomic tempo, while those in their 30-40s and women the humane tempo. In the second round of the survey, Mendelssohn was the only piece with a slow tempo. The phrases of its metronomic-tempo rendition felt regular without being prolonged; the performance in the humane tempo gave the contrasting impressions of being relaxed and moving forward.

Graph 3: Percentage of those in their 20s and 30-40s who chose the humane tempo in the first and second rounds of the survey



Graph 4: Percentage of men and women who chose the humane tempo in the first and second rounds of the survey



3. WRITTEN SURVEY AND ITS RESULTS

After listening to the music and choosing what they preferred, the audience was asked to partake in a written survey. Presented here are the combined results from the first and second rounds of the survey with 106 respondents in total (i.e. 36 in the first round and 70 in the second). Summarized below are the survey questions and answers in descending order of percentage:

Two questions were asked to find out when the audience member felt bored during a concert. The first question was: "When do you find the performer's interpretation of the piece uninteresting?" Answers as "When he/she plays without any change in the tone of sounds" and "When he/she plays without any change in dynamics" was marked by 29% of the audience, followed by "When the song is too long without any change in the tempo" (24%). The second question read: "Did you find the performer's way of playing or the piece itself uninteresting?" Answers included: "The piece itself felt boring" (34%), "The performer's bland performance was the problem" (30%), and "It seems like the performers do not communicate with each other when playing" (23%).

Asked when they found a concert interesting and touching, 47% of the respondents chose "when they feel the performers fully express the emotions and character of the pieces," showing that the audience likes it when the performers' feelings are effectively conveyed to them. Other answers included "when the musical pieces themselves sound interesting and touching" (24%) and "when sounds from the performers change dynamically" (19%).

After listening to music played in both metronomic and humane tempos, the respondents gave the most votes to the statement, "The humane tempo feels more interesting and comfortable as it seems to better express the performer's feelings" (55%), followed by "The metronomic tempo feels precise but not interesting or touching" (19%), and "The metronomic tempo is easier to listen to as it feels precise" (18%). What is noteworthy here is that some of the audience chose the metronomic tempo in 75-100% of the instances when listening to music and determining what they preferred, while they also picked the answer, "The humane tempo feels more interesting and comfortable as it seems to better express the performer's feelings," in the written survey. In the same vein, other respondents gave more votes to the humane tempo when listening to the music, but they replied, "The metronomic tempo is easier to listen to as it feels precise," in the written survey. In particular, 13 out of 48 professional musicians showed a gap (conflicting answers) between theoretical assumptions and their actual preferences determined after listening to the music.

4. OBSESSIONS WITH AND PERCEIVED NOTIONS ABOUT THE METRONOME AS ILLUSTRATED IN THE WRITTEN SURVEY RESULTS

Next are the answers from general public and music connoisseurs with experiences of playing musical instruments as well as from professional musicians; these represent the obsessions and perceived notions they have when practicing their instruments.

The first question was: "How do you set the tempo for a piece?" Of those surveyed, 37% said they "listen to CD or watch performances posted on YouTube before making the decision," followed by "checking the metronome markings on the score" (35%), and "listening to the teacher's opinions" (28%). A handful of other respondents said they "set a tempo that they find natural."

It is hard to break away from the traditional methods of learning a work of classical music, which has been repeated over a very long period of time, but we need to question how valuable it is to aim at someone else's tempo—instead of one's own—and continue to practice to produce the same results.-this subject requires another in-depth study.

Asked what they feel when practicing with a metronome, respondents said they "feel they have no sense of rhythm" (35%), "feel frustrated, angry, and irritated" (28%), "think it is only natural that it does not work out" (20%), and "feel happy as it works out" (15%).

When it came to the question, "What if your playing does not work well with the metronome?" the largest number of respondents (38%) said they would "practice until it works well," while 25% each replied they would "simply take as a reference how fast the metronomic tempo is" and would "try a bit and turn the metronome off once the tempo is roughly reached."

The next question was: "How strictly do you think you should follow the composer's own metronome markings?" Most of the respondents said they "try to follow the markings throughout the performance" (47%), while others explained that they "strictly follow the markings in the beginning but go with the flow once the performance begins" (27%) and "simply take the markings as a reference" (22%), hinting that a large number of those surveyed think the metronomic tempo should be followed.

Asked if they had ever practiced or played with a different tempo from the metronome marking on a score, a majority of respondents (65%) replied that they "do not care even if the tempo differs from what is mentioned on the score as long as they like the result." Two other answers, "It feels great to play freely without being bound by the tempo stated on the score" and "Failing to play with the stated tempo shows my limitations, and thus I feel bad" received 11% of the votes each. Noteworthy here is that many actually admit taking composer's tempo markings lightly. Of

those surveyed, 51% replied that they tried to follow the tempo markings throughout the performance, compared to only 36% who actually listened to the performance and said they liked the metronomic tempo better. This also hints at the discrepancy between theory and actual feelings.

Then came the question: "Pieces from which era do you think should work well with a metronome?" Of the respondents, 45% chose the Baroque period, followed by the Classical period (41%), the Modern period (14%), and the Romantic period (0%). Asked why they gave such answers, a vast majority of the respondents (72%) said that they thought so because "the temperance of expression was important in that era." However, in the case of 48 professional musicians this differed significantly after they had listened to music. Only 38% of professional musicians chose the metronomic tempo for a Bach piece played by a cellist (from the Baroque period), 38% for a Haydn piece (from the Classical period), 45% for a Mendelssohn piece (from the Romantic period), and a mere 25% for a Shostakovich piece (from the Modern period). This indicates a huge discrepancy between theory and practice as well as the prevalent perceived notions of different eras.

5. PERFORMER'S PERSPECTIVE

As the survey results suggest, a majority of respondents were frustrated when their tempo did not match the metronome. Many performers are trained to practice with a metronome for countless hours so they can perfect their technical facility. Whenever departing from the metronomic tempo, they would blame themselves for their lack of rhythmic sense, feeling unconfident and even anxious. When ensemble does not work, the performers typically conclude that it deviated from the metronomic tempo. If performers share their respective views and reach consensus through enough rehearsals, it would not matter at all if the performance does not follow the metronomic tempo. Expressing the flow of emotions naturally—rather than rushing in the mechanical metronomic tempo until they lose their breath—helps create better ensemble that the audience can sympathize with.

The humane tempo respects the composer's tempo markings yet reflects the unique flow of individual performers. Each performer has his or her own flow, making the performance all the more special. Performances become alive. Even the same performer cannot replicate the performance every time, as his or her performance will vary by the circumstances, location, and audience. The performer will change and evolve each time by timing (i.e. extent of preparation), audience (e.g. age, gender, general public/connoisseur/professional musician), location (i.e. concert hall, chamber hall, non-concert hall), and objective (e.g. audition, recital, ensemble). Against this backdrop, it is wiser to practice in wide-ranging tempo variations to be able to adapt quickly to different performing environments rather than to practice consistently with a single tempo as the goal.

6. CONCLUSION

All human performances are bound to have a humane tempo: When Bach's Suite is played by 100 performers, there will be 100 different performances. Even when the same person plays the same piece, there can never be exactly identical performances. With this in mind, performers are recommended to channel the energy they spend in practicing with a metronome—trying to play cookie-cutter songs just to achieve a precise tempo throughout—into searching for a humane tempo of their own that can make the true character of pieces stand out.

Performances may also be viewed as interpersonal dialogues, sympathetic expressions, and mutual reactions. Reactions are clearly different when a counterpart gets angry and yells, and when he or she talks affectionately. Most of us respond according to our counterpart's way of talking, which means we sympathize with and respond to how he or she says. Music is a form of art that expresses human emotions; sadness, joy, happiness, anger, etc. Music as a flow of emotions does not match well with metronome, which is a machine that cannot understand these complex emotions. Listening to music played in the metronomic tempo feels, often, suffocating. The performers and the audiences have trouble finding where to breathe, which affects both our comfort and concentration.

Also, listening to a rather mechanical playing, one can even predict what will happen next. The element of surprise is an essential part of a successful performance. Seen from the audiences' perspective, they would prefer a performance which is not totally predictable, but one that goes well beyond their expectations.

Human beings can reportedly concentrate on the same sound for only seven to ten seconds.26 This would mean that new stimulation or changes should be presented every seven to ten seconds in the concert as well. The colors and dynamics of sound as well as the visuals should be changed constantly to make the performance colorful and vibrant, so the audience can concentrate more easily on the performance; the tempo is no exception here. Reflecting human emotions, the humane tempo incorporates subtle but essential changes that cannot be found in the metronomic tempo, and thus can speak better to the audience.

7. FURTHER STUDY

The survey results may vary by the style of different performers, so a bigger number of sample performances need to be studied.

How the extent of training with a metronome and the metronome's influence on the performer since their childhood affect the audience's preferences should be examined.

The relationship between different populations of different languages and tempo preferences should also be explored to see if spoken languages have influence on musical preferences.

8. REFERENCES

Cohen, Patricia (9.26.2013). A new survey finds a drop in arts attendance. New York Times
Hocking, Rachel and Richard Letts (1.22.2008). International Classical Music Audience Attendance Trends 1990-2005
Medina, John (2014). Brain Rules
Weinschenk, Susan (2012). 100 Things Every Presenter Needs To Know About People.
Whitwell, David (2014). Essays on the Origins of Western Music. Essay Nr 65 Thoughts on Tempi www.whitwellessays.com/docs/DOC_961.doc

Session 8

[8A] Musical Practice

26 Susan Weinschenk, 『*100 Things Every Presenter Needs To Know About People*』 5.14.2012

SEOUL 17:00- 18:30 Thursday 7 Aug 2014

WHEN PRACTICE MAKES PERFECT: THE INFLUENCE OF DELIBERATE PRACTICE ON SKILL ACQUISITION IN MUSICIANS. A META-ANALYSIS

Friedrich Platz[1,2], Reinhard Kopiez[1], Andreas C. Lehmann[3], Anna Wolf[1]
[1]Hanover University of Music, Drama and Media, Germany, [2]University of Music and Performing Arts, Stuttgart, Germany, [3]University of Music, Würzburg, Germany
Time: 17:00

1. Background: Deliberate practice (DP) accumulated within the span of 10 years has been identified as a key concept in the skill acquisition of expert musicians (Ericsson, Krampe, & Tesch-Römer, 1993). However, despite the relevance of DP for the current discussion on the role of various factors influencing the development of expert performance (Hambrick et al., 2013), no meta-analysis (Cooper, Hedges, & Valentine, 2009) has been conducted to identify its empirical relevance in expert music performance by means of a quantifiable effect size. Moreover, considering the importance of meta-analyses (Kopiez, 2012), a systematic research synthesis might generally lead to new models of musical development, which could explain peak performance in music.

2. Aims: The aim of the study is the identification of an effect size, which quantifies the influence of deliberate practice on the acquisition of skills in the domain of music.

3. Method: We conducted a meta-analysis in five steps (Cooper, 2010): first, we started with an extensive and systematic review of the literature on DP in the domain of music performance – including peer reviewed journals as well as "grey literature" (Rothstein & Hopewell, 2009). This review resulted in about 15 relevant studies with data quality sufficient for a meta-analysis. Second, we re-analyzed every study's effect size for DP on performers' skill acquisition. Subsequently in a third step, we calculated the weighted average effect size of all aggregated studies. Fourth, the aggregated effect size of DP was tested for publication bias and reliability (Rothstein, Sutton, & Borenstein, 2005). Finally, we tested the validity of findings and the influence of external factors, such as family background, on DP by means of a meta-regression approach (Cooper et al., 2009).

4. Results: The meta-analysis is in the process of being completed, and the number of data sets included is still growing. However, we can currently see a significant effect size of $r > .50$ for DP.

5. Conclusions: Based on our results and in contrast to Hambrick et al.'s (2013) review of only a limited number of feasible experimental studies, we argue for the central role of DP in the process of skill acquisition in musicians.

AN ANALYSIS OF SELF-EVALUATION REPORTS OF THE PRACTICE PROCESS OF UNIVERSITY STUDENTS WHO ARE PIANO BEGINNERS

Towaki Shoshi[1], Yoko Ogawa[2]
[1]Graduate School of Education, Okayama University, Japan, [2]Okayama University, Japan
Time: 17:30

The purpose of this study is to investigate how university students who are piano beginners can improve by using self-evaluation reports. Over the past few decades, a considerable number of studies have been conducted on student learning processes. In music education, it is recognized students who learn to play musical instruments need to become autonomous learners (McPherson & Zimmerman, 2002). In addition, in education, self-evaluation reports have been used to understand student learning processes. There is research which shows that self-evaluation reports are effective for university students. However, very few attempts have been made to encourage Japanese university students who are piano beginners to use evaluation reports. This research investigated whether university students who are piano beginners would perform better or not if they used self-evaluation reports. Seventy-three Okayama University undergraduate students participated in the experiment. The subjects were randomly divided into an experimental group and a control group, and both groups took a pre-test, lessons and a post-test. The test had six parts: using a music sheet the students saw for the first time, they had to sing, play and beat the rhythm for two pieces of music (simple and difficult). After completion of the pre-test, just the experimental group were given a self-evaluation report and they filled in their self-evaluation score. In addition, students were told to write down their target score for the following test and create a practice plan for two months. Results showed the scores of the experimental group increased significantly ($p < .01$). One result shows the pre-test and post-test scores of the experimental group for singing (pre-test ave=1.83±0.98, post-test ave=4.46±1.86). Moreover, the average scores for self-evaluation of in the experimental group increased significantly ($p < .01$). In addition, there was a significant difference in the between the pre-test and the target score ($p < .01$). The experimental group set their target scores higher than their pre-test scores. It is inferred that this group practiced on the basis of this high goal setting. It was found from the results that self-evaluation reports are useful to help university students maintain their motivation to learn.

A STUDY ABOUT FLOW THEORY AND TEENAGER'S MUSICAL PRACTICE

Rosane Cardoso De Araujo[1], Margaret Amaral De Andrade[2]
[1]Federal University of Parana (ufpr)/ Cnpq, Brazil, [2]State University of Parana (unespar - Embap), Brazil
Time: 18:00

ABSTRACT

The focus of this study is on Flow Theory and teenagers' musical practice. For Czikszentmihalyi the Flow state enables the individual to fulfill an activity with great concentration and emotion. In the Flow state, the individual is totally immersed in the performance of an activity that provides a challenge reaching the threshold of his capability of control. The general objective of this research was to investigate the elements described by Csikszentmihalyi that compose the Flow experience in the musical practice of young musicians, specifically motivation to the musical practice, emotion, goals, concentration, perception of the competence/confidence. A survey was conducted with 35 students of music, aged 12-18 years, in 2012. The first category analyzed was motivation concerning musical practice and was possible to verify that the results were higher when

associated with the chosen repertoire, which means that repertoire is a decisive factor in motivation. Similarly, "repertoire" appears in the category feelings as an element of impact, since the levels of feeling (happiness), once associated to repertoire, were higher. Concentration and the use of goals were found. Most adolescents showed themselves aware of the need to organize their studies, and we can relate these results to the positive levels of concentration indicated. The category Self-confidence/ Competence had high levels for most teenagers. This category can point to an autotelic personality. The data observed in this study revealed that the students who participated in the research maintain in their daily lives many components that, according to Czikszentmihalyi (1999), can generate Flow. This study brings results that have implications for researches on motivation, cognition and music, more specifically for educators working with young students, because showed: a) the necessity of paying attention to the choice of repertoire when teaching a student, considering that it is an essential element for motivation and feeling; b) the need to organize a study routine, so as to enhance concentration; c) the need to strengthen the feeling of self-confidence and competence in the student, through the incentive of searching for goals that are possible to be reached.

1. BACKGROUND

This paper focuses on teenagers' musical practice in relation to Flow Theory. For Czikszentmihalyi (2003, 1999, 1996, 1992, 1990) the Flow state enables the individual to fulfill an activity with great concentration and emotion. In the Flow state, the individual is totally immersed in the performance of an activity that provides a challenge reaching the threshold of his capability of control. Flow Theory has been especially used in studies that focus on the analysis of the quality of the subjects' involvement in different activity. For Czikszentmihalyi (1999) the Flow State enables the individual to fulfill an activity with great concentration and emotion. This experience occurs through affective components of motivation that manage the performance.

In the context of musical practice, the Flow state occurs when the participation of an individual in a musical activity provides a positive feedback, increasing their sense of competence. This experience can be considered part of a process that favors individual persistence in the activity and it is created by contents that maintain a strong motivation in the individual. Emotions, goals and cognitive processes like concentration are included in these motivational processes. Researches on flow theory and music, in turn, have been undertaken by different authors as, for example, Custodero (2005; 2006), Addessi & Pachet (2007), Araujo & Andrade (2011), and Araujo et alii (2010).

2. AIM

The general objective of this research was to investigate the elements described by Csikszentmihalyi (1999) that compose the Flow experience in the musical practice of young musicians, specifically motivation to the musical practice, emotion, goals, concentration, and perception of the competence/confidence.

3. METHOD

A survey was conducted with 35 students of music, aged 12-18 years, in 2012. This research has three phases: a) elaboration of the questionnaire using data from two case studies conducted in previous research; b) validation of the questionnaire through a pilot study with 12 musicians; and c) a final survey with teenage musicians. The preparation of the questionnaire was made from a previous study: two case studies conducted with two musicians. The data collection in the case studies was made via semi-structured interviews, organized in topics: (1) characterization of the respondents: age, musical background, current activities; (2) daily musical practice: organization of practice, strategies, choice of repertoire, weekly study time, public performances; (3) motivational elements: preferably repertoire impressions on the group practice and individual practice, emotion, performance evaluation. Interview data were recorded and transcribed. To organize the aspects relating to the Flow components in the participants' musical practice, the data were categorized within five groups that emerged after the previous transcription analysis of the interviews. These categories were chosen based on the interview data and on their relation to the contents described by Czikszentmihalyi as components of the Flow Experience.

In the second phase, the categories that were observed in the analysis of the case studies were used to design a questionnaire that was tested with a sampling of 12 musicians in 2012. This test was essential to verify if this instrument of the data collection was satisfactory. After this test, the questionnaire was applied with teenage musicians.

The final questionnaire was designed in two parts: (1) collecting data about the research participants and (2) Likert scale about the categories: a) motivation and repertoire contribution; b) concentration during the study; c) Perception of the competence and confidence; d) Goals and strategies; and e) emotion.

4. RESULTS

Participants in this research were adolescents (n=35) aged between 12 and 18, regularly enrolled in a music school in the city of Curitiba (South of Brazil), studying different musical instruments.

Regarding their daily study routines, 25 participants (71%) affirmed that they "yes" maintain a study routine. 10 students (57%) stated that they "sometimes" maintain a routine. Zero participants declared not having a study routine. These results mean that all the participants held studying as a daily objective (goal) which, for the experience of Flow, is an important indicator (Csikszentmihalyi, 1999).

On being questioned as to the amount of hours dedicated per week to practicing, 12 students declared "until 03 hours per week" (34%); 13 students declared "until 05 hours per week" (37%); 06 students declared "until 10 hours per week" (02%), and 03 students declared "more than 10 hours per week" (01%). One student did not respond this question.

The Likert Scale applied in order to verify the teenagers' experience with elements that promote Flow was organized into five topics (*always, nearly always, sometimes, rarely* and *never*). The scale was divided into 10 questions, two for each category of analysis: *Motivation; concentration and "losing track of time"; self-confidence and competence; goals and strategies; emotion feelings.*

In the question about the adolescents' motivation while studying their instruments, the majority of them (n=33 students = 94%) indicated "always" or "nearly always" being motivated. None replied being "rarely" or "never" motivated. (see figure 1.)

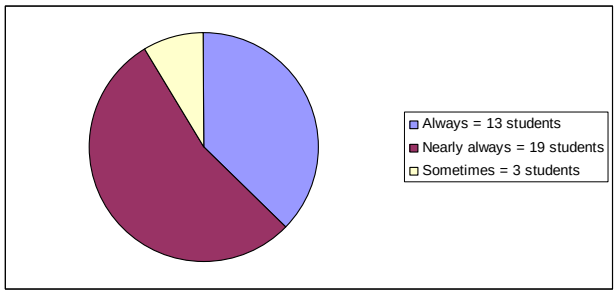


Figure 1: Motivation on studying an instrument

In this category about motivation concerning musical practice was possible to verify that the results were higher when associated with the chosen repertoire, which means that repertoire is a decisive factor in motivation.

Similarly, "repertoire" appears in the category feelings as an element of impact, since the levels of feeling (happiness), once associated to repertoire, were higher. Feelings of pleasure/joy in playing their instruments were indicated by 31 students (88%). When asked about the feelings related to playing a repertoire that they enjoyed, 32 students (92%) stated that they "always" (n=30 students) and "nearly always" (n=2 students) feel happy about this experience. (see figure 3)

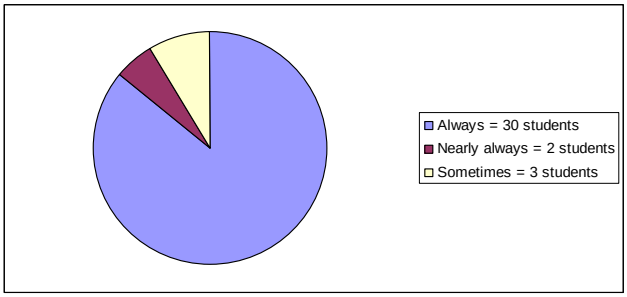


Figure 3 - Happiness while playing the repertoire of preference

According to the results, the subjects presented a greater motivation when specific repertoires were indicated. Repertoire may be considered a challenge factor, which means that it can engender motivation. Reeve (2006) states that ideally placed challenges, that is, those that instigate people to achieve certain tasks in conformity with their capacities, generate Flow as well as positive feedback, that is, the perception of one's own competence. Reeve also explains that an effort receiving no feedback loses emotional quality. By Czikszentmihalyi (1999), the existence of challenges compatible with each individual's capacity is something that promotes concentration and emotion in the performance of the activity, since they raise the levels of performance, energy and effort.

Czikszentmihalyi (1999) argues that positive emotions are inner states of consciousness that allow psychic energy to flow freely. Positive emotions such as happiness, strength or alertness, are states of "psychic negentropy". "Psychic negentropy" is a form of entropy in which the individual does not entertain self-reflectivity, maintaining the mind focused on the activity being performed. This process promotes the experience of Flow.

About concentration while studying, 33 students (94%) indicated that they were "always" and “nearly always” concentrated (see figure 2).

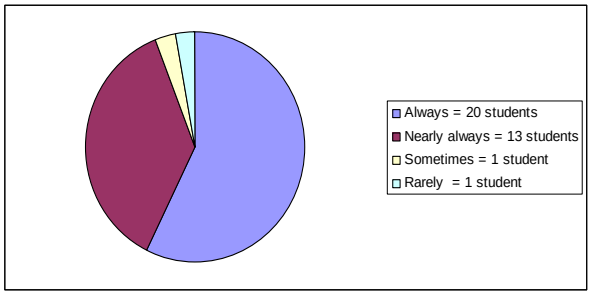


Figure 2: Concentration while studying

When questioned on the experience described by Czikszentmihalyi (1999) as "losing of track of time" (question 4), 23 students (66%) indicated having this experience "always" and "nearly always"; 8 students (23%) affirmed experiencing it "sometimes", and 4 students (11%) answered "rarely". None of the students have "never" experienced it.

29 participants (82%) affirmed they "always" or "nearly always" create goals so as to organize their studies. 32 students (91%) stated that they "always" or "nearly always" use strategies (for example: studying slowly the difficult parts; doing rhythmic reading, using of the exercises to applied technique; studying scales, etc)

Czikszentmihalyi (1999, 1992) explains that one of the necessary aspects for Flow to occur is the establishment of goals that aid in the process of concentration and organization of the given activity. According to Reeve (2006), goals generate motivation. In musical practice, goals and strategies can also be considered fundamental elements for self-regulation in learning. Self-regulation, as Bandura (1997, 2008) states, is the

individual's capacity for controlling one's own learning. Most adolescents showed themselves aware of the need to organize their studies, and we can relate these results to the positive levels of concentration indicated.

The category Self-confidence/ Competence had high levels for most teenagers. This category can point to an autotelic personality. Czikszentmihalyi (1999) states that when an individual is capable of performing different activities with enthusiasm, commitment and self-confidence, it is possible to assert that they possess an autotelic personality, meaning that they will execute activities for their own sake and the pleasure they may bring, having the experience itself as a main objective (and not other external factors).

5. CONCLUSIONS

The data observed in this study revealed that the students who participated in the research maintain in their daily lives many components that, according to Czikszentmihalyi (1999), can generate Flow. This study brings results that have implications for researches on motivation, cognition and music, more specifically for educators working with young students, because showed: a) the need to organize a study routine, so as to enhance concentration; b) the need to strengthen the feeling of self-confidence and competence in the student, through the incentive of searching for goals that are possible to be reached; c) the necessity of paying attention to the choice of repertoire when teaching a student, considering that it is an essential element for motivation and feeling.

6. REFERENCES

Addessi A.R.; Pachet F. (2007) Sistemas musicais interativos-reflexivos para educação musical. Cognição e Artes Musicais, *v. 2, n. 1, [pp. 62-72]*.

Araujo, R. C.; Andrade, M. A. (2011). Experiência de fluxo e prática instrumental: dois estudos de caso. Revista DAPesquisa, Volume 8, *[pp.553-563]*.

Araujo, R. C.; Cavalcanti, C. P.; Figueiredo, E. (2010). Motivação para prática musical no ensino superior: três possibilidades de abordagens discursivas. Revista da ABEM, 24, *[pp. 34-44]*.

Bandura, A. (1997). Self-efficacy: *the exercise of control. New York: Freeman.*

Bandura, A. (2008). A evolução da teoria social cognitiva. In: A. Bandura, R.G. Azzi & A. Bandura, R.G. Azzi & S. Polydoro (Orgs.). Teoria social cognitiva *[pp. 43-67]. Porto Alegre: Artmed.*

Csikszentmihalyi, M. (1990). Flow: *The psychology of Optimal Experience. New York: Harper & Row.*

Csikszentmihalyi, M. (1992). A psicologia da felicidade. *São Paulo: Saraiva.*

Csikszentmihalyi, M. (1996). Creativity: *Flow and the Psychology of Discovery and Invention. New York: Harper Collins.*

Csikszentmihalyi, M. (1999). A descoberta do fluxo. Psicologia do envolvimento com a vida cotidiana. *Rio de Janeiro: Rocco.*

Csikszentmihalyi, M. (2003) Good Business: *Flow, Leadership and Making of Meaning. New York: Viking.*

Custodero, L. A. (2006). Buscando desafios, encontrando habilidades: a experiência de fluxo e a educação musical. In: Ilari, B. (Ed.). Em busca da mente musical *[pp. 381-399]. Curitiba: Editora da UFPR.*

Custodero, L. A.(2005). Observable Indicators of Flow Experience: A Developmental Perspective on Musical Engagement in Young Children from Infancy to School Age. Music Education Research, *July, 07, n2, [pp.185-209]*

Reeve, J. (2006) Motivação e Emoção. *4.ed. Rio de janeiro: LTC.*

[8B] Music and Memory 2

KYOTO 17:00- 18:30 Thursday 7 Aug 2014

LONG TERM MEMORY FOR MUSICAL FORM: A NATURALISTIC INVESTIGATION USING POPULAR SONGS

Richard Ashley[1]
[1]*Northwestern University, USA*

Time: 17:00

1. Background: Psychological investigations into musical form have raised doubts and controversies as to its cognitive reality or significance. Taken together, these studies indicate at best a modest level of comprehension of musical form on the part of listeners. This study takes a different and more naturalistic approach to these questions, examining the accuracy of listeners' long-term memory for the form of popular songs acquired in the course of everyday life.

2. Aims: Popular songs, the everyday musical fare of most listeners, are used as our research materials. We hypothesize that listeners will have a high degree of veridical memory for the structure of popular songs, based not only on hierarchic tonal structures but also other parameters (lyrics, rhythm, melody, and timbre).

3. Method: Participants to date are persons with formal musical training; we will do an intergroup comparison with listeners without formal training. We use a recall task, in which participants produce verbal or graphic representations of the form of popular songs they know well only as listeners. These reproductions are scored against the original recordings for accuracy. Errors in reproductions give clues as to the aspects of popular songs leading to accurate encoding, especially with respect to "formal function" or "openness" of segments.

4. Results: Using correct recall of the order of formal segments as the basis for scoring responses, listeners are much more accurate than might be expected from earlier studies, with average accuracy of 73%. Mistakes tend to occur where adjacent formal segments lack formal or perceptual contrast with one another, although the hierarchical level of such lack of contrast varies from recall to recall. The high accuracy seems related to encoding by formally-meaningful ones such as "verse" or "prechorus" as opposed to smaller units, such as 2- or 4-bar segments.

5. Conclusions: Listeners can, and do, form quite veridical long-term memories for the forms of pieces of music. The ability of listeners to comprehend musical form is proposed to depend not only on familiarity (degree of exposure) but also on the nature of musical units used as "building blocks" and on the relationship of adjacent formal elements to one another.

IS THERE A MUSIC-STIMULATED ENHANCEMENT EFFECT ON AUTOBIOGRAPHICAL MEMORY? – TOWARDS SCHULKIND ET AL.'S (1999) HYPOTHESIS OF A REMINISCENCE BUMP

Reinhard Kopiez[1], Friedrich Platz[2], Anna Wolf[1]
[1]Hanover Music Lab, University of Music, Drama and Media, Hanover, Germany, [2]University of Music and Performing Arts, Stuttgart, Germany
Time: 17:30

1. Background: According to Schulkind et al. (1999), popular songs heard between the ages of 15 and 24 are remembered particularly well and have a stronger relationship to autobiographical memories when compared with music from other phases of life. However, up until today the role of a song's affective features on memories has remained unexplained. Moreover, it is unclear whether Schulkind et al.'s (1999) finding of an age- and music-specific memory effect is reliable in a different media landscape 15 years later.

2. Aims: Study 1 (a replication) investigated 3 questions: Do songs heard between 25 and 24 years (a) contribute to a better recognition rate, (b) cause a better emotional rating, and (c) produce more autobiographical memories than do songs heard at other ages? In Study 2 we went beyond Schulkind et al.'s (1999) study to investigate the influence of the affective characteristics of the music (valence/arousal) on the frequency of memories (see Eschrich et al. 2008).

3. Method: In Study 1 a group of adults (N = 66, mean age = 65 years) listened to a random sequence of number one songs on the charts (segment duration = 20 s) from the last 80 years. In an open answer format, participants gave written self-reports on memories related to the particular song. Answers were coded into "general" and "specific" memories. In Study 2 a group of adults matched for age with the participants of Study 1 (N = 22, mean age = 66 years) rated the same songs' affective characteristics on 6-point Likert scales in a 2-dimensional emotional space (valence/arousal).

4. Results: In general, Study 1 confirmed Schulkind et al.'s (1999) finding of an age- and music- specific memory enhancement effect. However, the effect size was small (general memories: r = .07; specific memories: r = .09). Study 2 revealed that memory enhancing songs were characterized by higher ratings for arousal (e.g., happy, angry) - but nor for valence ($\omega = 2.06$).

5. Conclusions: The overall memory enhancement effect for songs heard between the ages of 15 and 24 is small and should not be overemphasized in social work with the elderly.

PROCESSING ADVANTAGES FOR VOCAL MELODIES: ARE CHILDREN LIKE ADULTS?

Michael W. Weiss[1], E. Glenn Schellenberg[1], Sandra E. Trehub[1], Emily J. Dawber[1]
[1]University of Toronto, Mississauga, Canada
Time: 18:00

1. Background: Music cognition is typically studied with stimuli that have limited biological significance (e.g., instrumental sounds). We demonstrated, however, that adults remember melodies better when they are presented vocally (sung to la la la) rather than instrumentally (Weiss et al., 2012). Vocal melodies may have even greater processing implications for children, but we know little about children's processing of timbre, their memory for melodies, and how these abilities are affected by age-related cognitive change.

2. Aims: The present study aimed to ascertain the impact of vocal timbre on children's processing of melodies and to document age-related changes in recognition memory.

3. Method: In Experiment 1, 48 children ages 9-11 listened to 16 unfamiliar folk melodies (four each of voice, piano, banjo, or marimba), and rated their liking of each melody. Subsequently, they listened to the same melodies and 16 timbre-matched foils, and judged whether each melody was old or new. In Experiment 2, 5- to 6-year-olds (n = 40) and 7- to 8-year-olds (n = 40) were tested with a simplified design involving voice and piano timbres only.

4. Results: In Experiment 1, vocal melodies were recognized significantly better than instrumental melodies, which did not differ from one another. The vocal memory advantage was comparable across age, indicating that an adult-like voice advantage was evident in childhood. In Experiment 2, both age groups successfully differentiated old from new melodies, but memory was more accurate for the older group. Although 7- to 8-year-olds' differentiation of old from new melodies was significantly more accurate for vocal melodies than for piano melodies, that was not the case for younger children. Instead, younger children were significantly more likely to consider new melodies as old or familiar when they were presented vocally versus instrumentally.

5. Conclusions: The present study provides the first evidence of differential processing of vocal and instrumental melodies in childhood. For 5- and 6-year-olds, differential processing takes the form of a bias for vocal music such that unfamiliar vocal music seems more familiar than unfamiliar instrumental music. For children 7 and older, vocal melodies are recognized more readily, and memory becomes increasingly accurate.

[8C] Cross-Cultural Studies of Music

MONTREAL 17:00- 18:30 Thursday 7 Aug 2014

STATISTICAL UNIVERSALS REVEAL THE STRUCTURE AND FUNCTIONS OF HUMAN MUSIC

Patrick Savage[1], Steven Brown[2], Emi Sakai[1], Thomas Currie[3]
[1]Tokyo University of the Arts Department of Musicology, Japan, [2]Mcmaster University Department of Psychology, Neuroscience & Behaviour, Canada, [3]University of Exeter Centre of Ecology & Conservation, School of Biosciences, College of Life and Environmental Sciences, United Kingdom
Time: 17:00

1. .Background: Music has been called "the universal language of mankind". While contemporary theories of music evolution often invoke various musical universals, their existence is disputed by many skeptics and has never been empirically demonstrated.

2. Aims: Our goal was to provide the first quantitative estimates of the relative degrees of "universality" in a number of commonly cited candidate universals. We distinguish between two levels of universality: "absolute universals" that occur without exception and "statistical universals" that occur with some exceptions but significantly above chance. We also distinguish between two types of universals: "first-order" universals that concern presence or absence of individual features and "relational" universals that concern associations between multiple features.

3. Method: We manually coded the presence or absence of 32 features across the 304 recordings from the Garland Encyclopedia of World Music using pre-existing musical classification schemes. For candidate first-order universals, worldwide statistical significance was assessed using chi-squared tests, while for candidate relational universals, worldwide statistical significance was assessed using phylogenetic coevolutionary analysis based on language family classification. Regional consistency for both types of universals was assessed based on whether the direction of the trend was consistent across all of nine a priori defined geographic regions.

4. Results: Our analysis revealed no absolute universals, but strong support for 21 statistical universals that are consistent across all nine geographic regions sampled. These universals spanned not only acoustic features involving pitch, rhythm, form, performance style, and instrumentation, but also non-acoustic features of social context, such as dance, group structure, and gender. We also found moderate support for eight candidate universals that were consistent across most, but not all, geographic regions, while we found little or no support for a further 14 candidate universals.

5. Conclusions: Comparisons with human language and animal song suggest that some of these universals relate to basic vocal motor constraints, while others relate to cognitive and cultural factors linking music and sociality through a combination of mechanisms, including rhythmic entrainment, pitch matching, memory, arousal, and reward. The analyses reveal the common features that constrain the structure of human music, as well as the functions that music serves in human societies.

THE DENSMORE COLLECTION OF NATIVE AMERICAN SONGS: A NEW CORPUS FOR STUDIES OF THE EFFECTS OF GEOGRAPHY, LANGUAGE, AND SOCIAL FUNCTION ON FOLKSONG

Daniel Shanahan[1], Eva Shanahan[2]
[1]University of Virginia/louisiana State University, USA, [2]Independent Researcher, USA
Time: 17:30

ABSTRACT

The study of geographic effects on music can yield meaningful and important results related to the role of geography, language, and social functions on music. When examining possible such effects on music, however, there are a number of possible confounds presented by the datasets currently available. This paper examines these issues, and discusses the creation of a new dataset for studies of folk music, connections between music and language, and the possible effects of geography on musical style change.

1. BACKGROUND

Many studies examining the effects of both language and geography on music have employed the Essen Folksong Collection (Schaffrath, 1995), which contains roughly 6,000 European folksongs, and roughly 3,000 Chinese folksongs. For example, Aarden and Huron (2001) used the dataset, along with the encoded latitudinal and longitudinal information, to examine the differentiation in phrase types and mode usage throughout Europe. They concluded that Eastern European folksongs have more prototypical phrase endings (that is, they are more likely to end on the tonic), whereas Western European folksongs have more varied phrase endings. Additionally, they found a dominance of usage of major mode in Italy, and a dominance of minor in Eastern Europe. Interestingly, they found a greater use of the major mode that seemed to follow the Rhine River northward. They concluded that the influence of Italian music projected northward along the main river connection. This was consistent with the notion of the spreading influence of major-mode Italian practice into northern Europe along the Rhine.

The Essen dataset has also been used to examine linguistic effects of folksongs. Huron (1996) used the collection to demonstrate the prevalence of the "melodic arch" in folksongs, and Shanahan and Huron (2011) used the dataset to examine the possibility of intervallic "phrase compression" analogous to that present in speech.

1.1. The Limits of the Essen Folksong Collection

Unfortunately, there are a number of reasons why the Essen Folksong collection might not be an ideal dataset for such studies. First of all, it is collected from many different sources, including transcriptions from different periods, scholars, and locations. Secondly, the dataset is somewhat apocryphal: the specifics of each transcription (such as the date, the transcriber, the location, the specifics about the musicians) are not included. Often times very important facts are left out. For example, with the Chinese folksong dataset, it is unclear whether the transcriptions are of vocal or instrumental songs. The European folksongs are predominantly vocal, but it's likely (although not certain) that the Chinese folksongs are instrumental. Additionally, although, the Essen folksong collection provides a great deal of data, it is from a relatively small geographic region, all composed over the course of about 300 years.

1.2. What Would an Ideal Dataset Entail?

Ideally, a dataset that would be used to examine the effects of language and geography on music would fulfill a number of criteria. These include:

1. Being transcribed before the advent of mass media, to minimize the effect of the consolidation of disparate cultures and languages by mass media.
2. Being transcribed by a single individual, or a small group of individuals, in order to minimize transcription effects.
3. Covering a relatively large geographic area
4. Being collected over a relatively brief amount of time.

Although it seems like such a dataset might be difficult to find, one possible option might be found in the work of American ethnologist Frances Densmore (1867-1957).

1.3. THE DENSMORE COLLECTION

Densmore was employed by the Bureau for American Ethnology (BAE), beginning in 1907. Over the course of her career, she embarked on 79 field trips to 54 locations. She made around 3,500 recordings, transcribed more than 2,300 songs, and she published 16 books and hundreds of articles.

This project has consisted of the encoding of Densmore's transcriptions into searchable formats (such as **kern and MEI), and has focused on her books, which include the most information about the transcriptions, the context of the recordings, and the function of each specific song.

1.4. A Brief History of the Densmore Encoding Project

Empirical musicologists have been interested in encoding Densmore's collection of transcriptions for more than a decade. The project has largely consisted of encoding the transcriptions from each book into the kern format, for use with the Humdrum Toolkit (Huron, 1995). Paul von Hippel encoded excerpts of the first book of Chippewa songs in 1998. David Huron encoded the *Pawnee* and *Mandan* books in 2000, and Craig Sapp encoded the *Teton Sioux* book in 2002. Over the past year, the authors of the present study have encoded all of the remaining books.

Table 1: Frances Densmore's collections of transcriptions, and the number of pieces each contains.

Book	Number of Pieces	Year Published
Acoma, Isleta, Cochiti, and Zuni Pueblos	82	1957
British Columbia	98	1943
Cheyenne and Arapaho	72	1936
Chippewa (I)	249	Collected 1907-1909
Chippewa (II)	182	1913
Choctaw	71	1943
Maidu	53	
Mandan and Hidatsa	74	1923
Menominee	144	1932
Nootka and Quileute Music	132	1939
Northern Ute	116	1922
Papago	170	1929
Pawnee	100	1929
Seminole	247	1956
Teton Sioux	246	1918
Yuman and Yacqui	134	1932

Each encoding contains the information of each transcription, including the date of performance, the location, the social function (e.g. children's songs, hunting songs, etc.), whether the song was sung by a male or female, the linguistic group of the Native American tribe, whether or not it is typically considered a tone language, and the specific geographical location of the group. The corpus is now available in both **kern notation and MEI (Music Encoding Intiative) formats.

1.5. Some Issues with the Densmore Collection

Although this will hopefully prove to be a useful dataset for empirical musicologists, there are a number of issues that one must consider when using the Densmore corpus. Firstly, the corpus employs Western notation to depict music that does not necessarily conform to Western notational standards. As such, the transcriptions encoded might be viewed as approximations (as is all music notation). Secondly, one should consider the ethical implications of using such transcriptions. Densmore lacked formal training as an anthropologist, and her attitude toward her subjects in the early part of her career is often described as condescending and patronizing. Often times, the music being transcribed might not be meant to be performed or displayed outside of certain social situations by certain individuals. We as scholars must take such ethic considerations into account when employing such datasets. For more on such issues, see Huron's early description of the project, and the discussion of ethics (2002). The current project simply attempts to increase the accessibility of these transcriptions, in the hope of facilitating research pertaining to possible connections between music, geography and language.

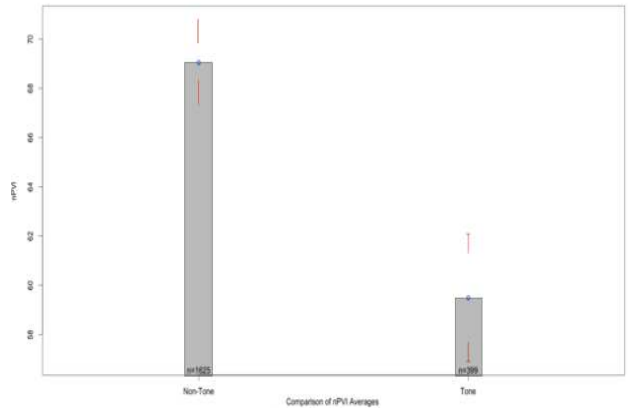


Figure 1: A comparison of normalized pairwise variability indices (nPVI) of melodies in groups that employ non-tone languages (on the left) and tone languages (on the right).

2. EXAMINING NPVI BETWEEN GROUPS

One example of a study that might be done with such a dataset might be a comparison of the normalized pairwise variability (nPVI) between groups that employ tone languages and those that do not. The nPVI is a metric ranging from 0 to 200 that analyzes the degree of variation in duration from one element to the next. A low value means that there is little variation in duration, while a high value means that the variation is quite high. Languages differ significantly in the nPVI. For example, Grabe and Low (2002) and Ramus (2002) found that the nPVI of Dutch and English tend to be much higher than Spanish or French. Patel and Daniele (2003) analyzed the instrumental melodies of both French and British composers, and found that there was a significant difference in nPVI between French and English melodies, corresponding with the languages. French had far less variation than English, and their melodies exhibited a similar disparity.

With the Densmore collection, we can examine the nPVI of the melodies of groups that primarily employ a tone language, compared with those that do not. As can be seen in Figure 1, groups that employ tone languages exhibit a much smaller nPVI. This might suggest that, while tone languages require specific pitch associations to convey meaning, non-tone languages might place more emphasis on rhythmic variability as a way of conveying meaning.

3. CONCLUSION

It is our hope that this corpus will provide a useful resource for scholars interested in music, language, geography, and social function. The dataset is available at www.github/humdrum-tools/densmore. Future work will employ the geographical tagging of each collection to look at the possibility of geographic effects on music.

4. ACKNOWLEDGMENTS

This work was made possible by the great work done by Paul von Hippel, David Huron, and Craig Sapp. Craig Sapp created the converter for MusicXML to kern, as well as kern to MEI, which has facilitated greater access of the dataset.

5. REFERENCES

Aarden, B., & Huron, D. (2001). Mapping European Folksong: Geographical Localization of Musical Features. Computing in Musicology, 12, *169-183.*

Densmore, F. (1910) Chippewa Music, Vol.1, *Bureau of American Ethnology.*

Densmore, F. (1913) Chippewa Music, Vol.2, *Bureau of American Ethnology.*

Densmore, F. (1918) Teton Sioux Music, *Bureau of American Ethnology.*

Densmore, F. (1922) Northern Ute Music, *Bureau of American Ethnology.*

Densmore, F. (1923) Mandan and Hidatsa Music, *Bureau of American Ethnology.*

Densmore, F. (1929) Pawnee Music, *Bureau of American Ethnology.*

Densmore, F. (1929) Papago Music, *Bureau of American Ethnology.*

Densmore, F. (1932) Yuman and Yacqui Music, *Bureau of American Ethnology.*

Densmore, F. (1936) Cheyenne and Arapaho Music, *Bureau of American Ethnology.*

Densmore, F. (1943) Nootka and Quileute Music, *Bureau of American Ethnology.*

Densmore, F. (1943) Music of the Indians of British Columbia, *Bureau of American Ethnology.*

Densmore, F. (1943) Choctaw Music, *Bureau of American Ethnology.*

Densmore, F. (1956) Seminole Music, *Bureau of American Ethnology.*

Grabe, E. and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. Laboratory Phonology 7, *edited by C. Gussenhoven and N. Warner Mouton de Gruyter, Berlin, pp. 515–546.*

Huron, D. (1995). The Humdrum Toolkit: Reference Manual. *Menlo Park, California: Center for Computer Assisted Research in the Humanities.*

Huron, D. (1996). The melodic arch in Western folksongs. Computing in Musicology, *Vol. 10, pp. 3-23.*

Huron, D. (2002) Ethics in The Densmore Project: Music of the Native Peoples of North America (www.musiccog.ohio-state.edu/Densmore/ethics.html), accessed on April 1, 2014.

Ramus, F. (2002). Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues. Annual Review of Language Acquisition, 2, *85–115.*

Schaffrath, H. (1995). The Essen Folksong Collection. D. Huron (Ed.), Stanford, California: Center for Computer Assisted Research in the Humanities.

Shanahan, D., & Huron, D. (2011) Interval Size and Phrase Position: A Comparison between German and Chinese Folksongs, Empirical Musicology Review, 6, *187–197.*

COGNITIVE, AFFECTIVE AND EMBODIED RESPONSES TO RHYTHMIC COMPLEXITY: A CROSS-CULTURAL COMPARISON OF WEST AFRICAN AND NORTH AMERICAN NON-MUSICIANS.

Maria Witek[1], Jingyi Liu[2], John Kuubeterzie[3], Senyo Adzei[4], Appiah Poku Yankera[5], Morten Kringelbach[6], Peter Vuust[1]
[1]Center of Functionally Integrative Neuroscience, University of Aarhus, Denmark, [2]Williams College, USA, [3]African Footprint International, Ghana, [4]Department of Music and Dance, University of Cape Coast, Ghana, [5]Department of Educational Foundation, University of Cape Coast, Ghana, [6]Department of Psychiatry, University of Oxford, United Kingdom

Time: 18:00

There is an ongoing discussion about the extent of difference in rhythm and metre perception between African and European music traditions. Some have claimed that African rhythms are not only structurally more complex, but the mechanisms of metre perception in African listeners are more sophisticated. However, such claims have been criticised as culturally essentialising and exoticising. Here, we report from a cross-cultural empirical study which addressed cognitive, affective and embodied responses to rhythmically complex patterns in two cultural groups of non-musicians: Ghana and North America. It was found that stability ratings of generic syncopated patterns did not differ significantly between groups, indicating that the metric frameworks of the two groups were comparable. However, subjective ratings of desire to move and feelings of pleasure in response to groove-based rhythms were different. Participants from Ghana rated the grooves as more movement- and pleasure-

eliciting overall. Furthermore, while there was an inverted U-shaped relationship between degree of rhythmic complexity in grooves and ratings of wanting to move and experience of pleasure in the North American participants, this inverted U-shape was less pronounced in the Ghanaian group: there was no significant difference when the grooves had low compared to medium degree of rhythmic complexity, only when low and medium was compared with high. In other words, while North Americans prefer intermediate degrees of rhythmic complexity in groove, Ghanaians experience both low and intermediate degrees of complexity as equally pleasurable and corporeally entraining. Thus, our findings suggest that metre perception may be no different in West African compared to North American non-musicians, but culture affects the ways in which complexity influences affective and embodied responses to musical rhythm.

[8D] Singing 1

LIÈGE 17:00- 18:30 Thursday 7 Aug 2014

THE EFFECTS OF TIMBRE AND MUSICAL TRAINING ON VOCAL PITCH-MATCHING ACCURACY

Priyanka Shekar[1], Takako Fujioka[1]
[1]Center for Computer Research in Music and Acoustics, Stanford University, USA
Time: 17:00

ABSTRACT

Pitch control is one of the most essential skills when singing. Previous studies indicated that in poor-pitch singers, pitch-matching accuracy was greater when the timbre of the reference tone was vocal, as compared to non-vocal. However, it is not consistently shown as to whether this vocal timbre advantage exists across musically trained and untrained individuals. The present study investigated how vocal pitch-matching was influenced by the timbre of reference tones and the type of musical training experience. Pure tones and recorded voice were used as reference stimuli in a single-tone pitch-matching task. Vocalists, instrumentalists and non-musicians were tested. Acoustic analysis on the vocal responses revealed that the pitch accuracy was significantly better in response to voice across all groups, while vocalists performed not differently from instrumentalists. This suggests the universality and robustness of the vocal timbre advantage, which is not influenced by intense vocal practice. The pitch variability across trials was significantly larger in non-musicians than in the two musician groups. Non-musicians also showed larger heterogeneity across individuals.

1. INTRODUCTION

During vocal performance, pitch control is a core skill: one may be required to accurately match pitch to a reference tone, such as piano accompaniment or fellow choir members.

Previous studies have shown that musical experience greatly enhances vocal pitch matching accuracy, even if the training is not limited to vocal training. For example, instrumentalists (Amir, Amir, & Kishon-Rabin, 2003; Nikjeh, Lister, & Frisch, 2009) and vocalists (Estis, Dean-Claytor, Moore, & Rowell, 2011) have better pitch-matching accuracy than non-musicians, regardless of whether the reference stimulus was given in pure tone or complex tone. Interestingly, however, training experience may not be the sole factor in determining pitch-matching performance, because Watts, Moore, & McCaghren (2005) and Estis et al. (2011) found that non-musician subjects can be further classified into two groups: moderately good singers and poor singers who are largely inaccurate.

Several studies have attempted to characterize singing ability in non-musicians in detail (Hutchins & Peretz, 2012; Dalla Bella, Giguère & Peretz, 2007; Pfordresher, Brown, Meier, Belyk & Liotti, 2010), where poor singing cannot be simply attributed to a motor problem or perceptual problem. Interestingly, some studies revealed that in poor-pitch singers, pitch-matching accuracy was greater when the reference tone was presented in vocal timbre, as compared to non-vocal timbre (Watts & Hall, 2008; Hutchins & Peretz, 2012; Lévêque, Giovanni, & Schön, 2012; Granot, Israel-Kolatt, Gilboa & Kolatt, 2013). Because this advantage was greater when the vocal stimulus was similar in timbre to the subject's own voice (Hutchins & Peretz, 2012; Granot et al., 2013), the origin of this advantage may be due to an improved auditory-motor transformation process, where similar vocal cues trigger a motor command more effectively. At the same time, this advantage could also be the result of rich experience in hearing one's own voice.

If such vocal timbre advantage manifests from a privileged auditory-motor transformation process in hearing one's own voice, it could also exist universally, regardless of musical training, experience and ability. This question remains not fully conclusive, as previous studies have examined only a limited set of subject categories, and/or revealed inconsistent results. As introduced earlier, clear vocal timbre advantage was found in self-identified poor singers (Lévêque et al., 2012; Granot et al., 2013) and non-musicians (Watts & Hall, 2008; Hutchins & Peretz, 2012; Moore, Estis, Gordon-Hickey & Watts 2008), while other results were not so certain (Price, 2000). This inconsistency may be attributable to stimulus and task setup, and subject selection procedure. Hutchins & Peretz (2012) and Moore et al. (2008) similarly found an advantage in using own voice in female non-musician subjects. However, in the latter study, unlike the former, there was no distinction between neutral female voice and complex tones.

As for musically trained individuals, if the long-term experience in hearing one's own voice is the source of vocal timbre advantage, musicians with vocal training should exhibit greater accuracy and vocal timbre advantage than instrumentally trained musicians. Once again, research is sparse, yielding inconclusive results. In Estis et al. (2011) vocalists performed more accurately than non-vocalists, but it is unclear whether non-vocalists included instrumentally trained musicians. One study that explicitly compared vocalists and instrumentalists (Nikjeh et al., 2009), found no significant difference in the two groups when synthesized piano tones were used. Another study (Ogawa & Murao, 2004) observed an interesting result that vocalists sang flat when matching piano and pure tones, and sharp when matching vocal sounds.

If experience in hearing a similar timbre to one's own voice leads to vocal timbre advantage, we may expect that instrumentalists perform better in response to non-vocal sounds compared to vocalists and non-musicians (thus, diminished vocal timbre advantage). Another interesting insight from the study by Ogawa & Murao (2004) was the directionality of the pitch error: note that this was the only study that showed a significant difference in the directionality of pitch-matching accuracy between vocal sounds and pure tones, and all other studies mentioned here examined the pitch error measure as an absolute value of the deviation from the target pitch. Moreover, inconsistency across trials seems larger in

non-musicians than musicians in a few studies (Estis et al., 2011; Lévêque et al., 2012), but unfortunately solid results with statistical analysis have not been described to date.

Thus, the purpose of this study was to investigate in detail how pitch-matching accuracy, in both the directionality and magnitude of the pitch deviation, as well as the variability across trials, is influenced by the timbre of reference tones and musical training experience. We examined non-musicians, vocalists, and instrumentalists, using pure tones and vocal sounds as a reference. We analyzed the vocalization responses in three types of measures: signed and unsigned (absolute) pitch deviation from the target pitch, and the standard deviation across trials. We hypothesized that vocal timbre advantage may exist in all groups, and that musical training content may serve as an additional determinant of pitch-matching accuracy leading to vocal timbre advantage, such that musical training in voice and instruments may affect the performance differently.

2. PARTICIPANTS

30 participants (10 females, age 20-43 years) were recruited from Stanford University through an email advertisement specifically targeted at vocalists, instrumentalists and non-musicians. Vocalists and instrumentalists should have more than 5 and 7 years of formal training, respectively, and an active current practice status of more than 2 hours per week. Non-musicians should have less than 2 years of formal training. Recruited individuals were screened to confirm they had no history of hearing or vocal impediment, and grouped by self-identification, and according to a questionnaire. 8 participants were determined to be vocalists (age: M = 25.4 years, SD = 7.6), with considerable years of vocal training (training: M = 10.6 years, SD = 5.8). 12 participants were grouped as instrumentalists (age: M = 23.2 years, SD = 2.6), with considerable years of instrumental training (training: M = 13.0 years, SD = 3.6). 10 participants were grouped as non-musicians (age: M = 25.4 years, SD = 2.7), with little or no years of musical training (training: M = 0.5, SD = 0.7). Participants grouped as vocalists had often had some years of instrumental training (training: M = 9.5 years, SD = 7.9), but the rate of instrumentalists who had had some years of vocal training was much less (training: M = 0.9 years, SD = 2.2). In the case where participants met the criteria for both vocalists and instrumentalists, their self-identification was used for the group assignment. Two recruited individuals who did not meet grouping criteria, were excluded from the study. Participants were offered an honorarium in the form of a $10 gift voucher for their participation.

3. STIMULI AND APPARATUS

Target tones of pure and vocal timbres were used in testing. Target fundamental frequencies were determined separately for Alto (low female), Tenor (high male) and Bass (low male) voice types. 7 tone frequencies were randomly selected such that they existed in the central octave for each range (Kostka, Payne & Schindler, 2000), but did not evoke any sense of tonality (Table 1). Pure tones were produced with Audacity software (version 2.0.3, The Audacity Team). 7 vocal sounds at the same fundamental frequencies were produced by recording experienced vocalists singing scales in the same octave as above on the vowel /a/ ("as in ahh or father"). A very small degree of pitch shifting was then applied to the vocal sounds using an audio editing software Audacity, to exactly match their continuous pitches to the pure tones. Target tones were 1.0s in length, including a linear ramp of 0.025s applied at onset and offset. We did not use a soprano range because the recording obtained from a Soprano vocalist contained considerable vibrato that would interfere with the pitch-matching task performance. Target tones were played to participants over a pair of loudspeakers (M52, Klein + Hummel), at an approximate distance of 95 cm from the participant's head. Vocalization responses were recorded with a microphone (SM58, SHURE). The experimental regime was programmed with PsychoPy software (Peirce, 2007), and was presented to participants on the researcher's laptop computer (Apple MacBook Pro Retina).

Table 1: Target tone frequencies in Alto, Tenor, and Bass vocal ranges.

Ratio to the reference frequency (1 = unison, 2 = octave)	Fundamental Frequency (Hz)		
	Alto	Tenor	Bass
1.00 (reference)	246.9	164.8	103.8
1.16	381.2	285.5	190.6
1.43	353.2	235.7	148.5
1.50	371.4	247.9	156.2
1.76	435.7	290.8	183.2
1.81	448.1	299.1	188.4
1.93	476.3	317.9	200.3
1.97	649.7	486.7	324.8

4. PROCEDURE

Upon arrival, participants received a detailed explanation of the study, and gave informed consent for participation. Following an electronic screening questionnaire, a warm-up phase required the participant to determine a comfortable voice type (Alto, Tenor or Bass). If the participant did not know their voice range, they wer given the option to listen to a recording of a one-octave scale sung in each range by the same vocalists as used in the vocal stimuli, and then to select a suitably comfortable part. 10 participants selected the Alto voice (female), 6 selected Tenor, and 14 selected Bass.

In the test procedure, each trial cycle begins with a visual cue displayed on the screen for 1.5s, followed by presentation of the target tone of 1.0s length. The cue then switches to indicate recording mode and the participant is required to vocally reproduce the target tone within a 5.0s interval, as immediately and as faithfully as possible on the vowel /a/, and is automatically recorded. There was an option to skip ahead to the next trial if they had completed their response or were unable to, via key press. Between blocks, participants were given the control when to proceed with the next block via key press. Trials were performed for 7 target pitches in 2 timbres, comprising a mixed block of 14 target tones in pseudo-random order, such that no 2 identical pitches occurred consecutively. Each target tone was repeated across 3 blocks, thus there were 42 trials in total. Prior to the actual testing phase, a 5-trial practice block was given. The Stanford University Institutional Review Board approved this experiment.

5. DATA ANALYSIS

Vocalization response recordings were analyzed for pitch, using the Praat software (Boersma, 2002). The mean pitch value was extracted from the central 0.5s segment of the response, in an attempt to analyze only the stable portion. Due to a technical problem in recording, about 0.8% of trials across all responses were discarded. Pitch difference (error) from the target tone was then calculated for each trial. All pitch difference information was converted to a log-scale in cents (1 semitone = 100 cents) in order to make meaningful comparisons between different target frequencies. In a small portion of trials (5%), clear downward octave shifts were observed. When the pitch was closer than 150 cents to the octave below, 1200 cents was added for correction. Signed and unsigned pitch differences were separately evaluated in order to examine the directionality of pitch error. Pitch difference values falling outside 2 standard deviations for each participant for each condition were excluded as outliers at this stage. For mean and standard deviation, separate two-way mixed-design analysis of variance (ANOVA) were performed using a between-subject factor Group (non-musician, instrumentalist, vocalist) and a within-subject factor Timbre (pure tone, voice). An additional ANOVA was performed across vocalists and instrumentalists alone, for a more direct comparison between the two musician groups. Post-hoc tests were performed using two-sample t-tests for comparison between the groups, and pair-wise t-tests for comparison between the timbre conditions. Significance level was set at 0.05.

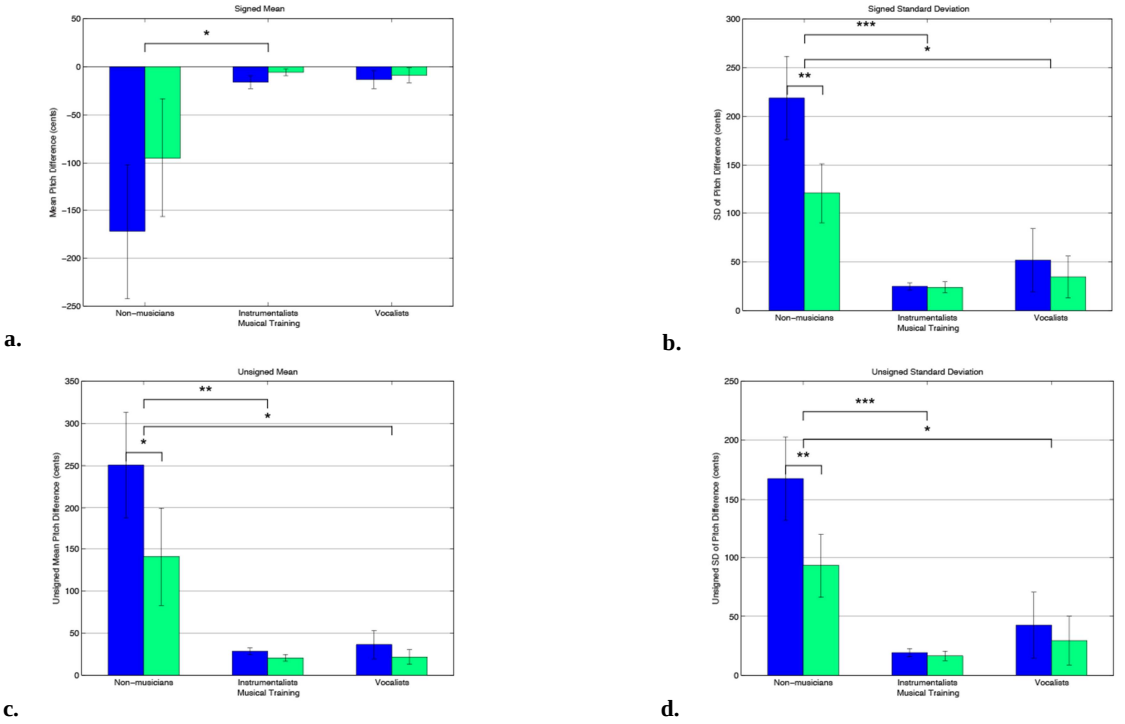


Figure 1: Pitch difference to the target frequency during pitch-matching in non-musicians, instrumentalists and vocalists – a. signed mean, b. unsigned mean, c. signed S.D., d. unsigned S.D. Error bars represent the SEM.
*** P < 0.05, ** P < 0.01, *** P < 0.001.**

6. RESULTS

For the signed pitch difference, all the mean values turned out to be negative, indicating the sung pitch was flat compared to the target fundamental frequency across all groups (Figure 1a), although the amount of error was strikingly larger in non-musicians compared to the two musician groups. The ANOVA revealed a significant main factor Group ($F(2,27) = 3.690$; $P = 0.0383$), confirming the effect of musical training. Post-hoc comparison showed significant difference between non-musicians and instrumentalists ($P < 0.05$) (Figure 1a). The differences between non-musicians and vocalists, and instrumentalists and vocalists, were not significant. Timbre as a main factor was significant ($F(1,27) = 5.231$; $P = 0.0302$). However, post-hoc tests revealed that the difference between pure tone and voice timbres was marginal within all groups – non-musicians (N.S.; $P = 0.0788$), instrumentalists ($P = 0.0968$), vocalists ($P = 0.0732$). Interaction Group × Timbre was also marginal (N.S.; $P = 0.0621$).

For standard deviation of the signed pitch difference (Figure 1b), the ANOVA showed a highly significant main effect of Group ($F(2,27) = 11.908$; $P = 0.0002$), due to the large variability across trials for non-musicians, compared to instrumentalists ($P < 0.0001$) and vocalists ($P < 0.05$). This clearly indicates that non-musicians not only sung largely flatted pitch, but also their variability across trials was remarkably large. In contrast, the difference between instrumentalists and vocalists was not significant. Timbre as a main factor was also highly significant ($F(1,27) = 12.581$; $P = 0.0014$), but within a group this was only significant in non-musicians ($P = 0.0097$) (Figure 1b). Interaction Group × Timbre was also significant ($F(2,27) = 8.067$; $P = 0.0018$) as the difference in variability between the two conditions was by contrast, rather small in the two musicians groups.

We also examined the unsigned pitch difference (i.e. absolute value of pitch difference) to be able to relate our results to the previous studies. The ANOVA for the mean value showed a significant main factor Group ($F(2,27) = 8.394$; $P = 0.0015$). The post-hoc comparison revealed significant difference between non-musicians and both instrumentalists ($P < 0.01$) and vocalists ($P < 0.05$) (Figure 1c). The differences between instrumentalists and vocalists were not significant. Main factor Timbre was significant ($F(1,27) = 11.933$; $P = 0.0018$). However, post-hoc tests

revealed that the difference between pure tone and voice timbres was only significant in non-musicians ($P < 0.05$). The interaction Group × Timbre was significant ($F(2,27) = 6.928$; $P = 0.0037$), because of this clear timbre difference in non-musicians.

In the standard deviation of the unsigned pitch difference (Figure 1d), the ANOVA revealed a highly significant main effect of Group ($F(2,27) = 9.154$; $P = 0.00092$), due to the large variability across trials for non-musicians, compared to instrumentalists ($P < 0.001$) and vocalists ($P < 0.05$). The difference between instrumentalists and vocalists was insignificant. Timbre as a main factor was also highly significant ($F(1,27) = 13.687$; $P = 0.00097$), but within a group this was only significant in non-musicians ($P = 0.0088$) (Figure 1d). Interaction Group × Timbre was also significant ($F(2,27) = 7.987$; $P = 0.0019$).

To test whether vocal timbre advantage existed in musician groups, supplementary ANOVA's considering only instrumentalists and vocalists were performed, yielding the following results: for signed pitch difference, none of Group, Timbre, or their interaction was significant. Only Timbre approached significance in the mean of the signed pitch difference (N.S.; $P = 0.0554$) showing slightly more flatted pitch sung in response to pure tone compared to voice. For unsigned pitch difference, main factor Timbre showed significance in the mean ($F(1,18) = 6.460$; $P = 0.0205$), and was marginal in the standard deviation (N.S.; $P = 0.0998$). Group and its interaction with Timbre were both insignificant.

7. DISCUSSION

Our results show that vocal timbre advantage was observed in non-musicians, in both signed and unsigned pitch difference measures for mean and variability. The source of this error and variability in non-musicians may be due to the presence of poor-pitch singers within the group. This result extends previous findings in both those which have used only the mean pitch difference (Watts & Hall, 2008; Hutchins & Peretz, 2012; Lévêque et al., 2012; Granot et al., 2013), and those which indicated the standard deviation as a observation measure but lacked rigorous statistical analysis (Estis et al., 2011; Lévêque et al., 2012). As the standard deviation measure here indicates how people are consistently able to reproduce a similar pitch across performances, this seems useful in elucidating an additional important characteristic that defines singing ability, such as accuracy and precision (Pfordresher et al., 2010).

The novel finding in our study is that all groups sang flat during reproduction regardless of the stimulus categories, and the vocal timbre advantage manifested as an even flatter pitch produced for pure tone stimuli. This is partly in line with Ogawa & Murao (2004)'s results for vocalists, showing the more flat sung pitch for non-vocal stimuli compared to vocal stimuli. However, their results showed slightly sharpened pitch for vocal stimuli, which was not the case in our data. This discrepancy may be due to our small sample size, target pitch selections, and different subject inclusion criteria. This result may also be related to the compressed sung interval shown in non-musicians regardless of their level of accuracy (Pfordresher et al., 2010). It is possible that our prescribed target pitch range might have been too high for most of the participants even though we used the ranges well defined in typical choral music practice. Nevertheless, the sung pitch may be generally flattened across individuals with different levels of pitch matching competency. This should be further examined in future research with individually adjusted vocal range.

There was a clear difference between non-musicians and the two musician groups, in agreement with previous studies demonstrating similar enhanced singing ability in musicians (Amir et al., 2003; Nikjeh et al., 2009; Watts et al., 2005). However, we did not find any difference between instrumentalists and vocalists. For the pure tone stimuli, vocalists may have been more flat and variable in performance than instrumentalists, there were no significant differences in our statistics. It is possible that there may be a difference in the underlying mechanism or strategy employed in the pitch-matching task between the two musician groups, though our measures were not sufficiently sensitive to capture it. Future research with larger trial size, sample size and task variety can delineate how different musical training experience can affect pitch production processes in more detail.

Though non-musicians were categorized based on musical training, and were not self-identified poor-pitch singers, the majority turned out to be largely inconsistent and incompetent compared to musicians. On the other hand, we also observed a few individuals whose performances were compatible with musically trained individuals. This observation is echoed by Estis et al. (2009), which subcategorized the untrained group as accurate and inaccurate, and further found that some untrained accurate individuals displayed pitch-matching accuracy similar to trained individuals. The source of poor-pitch singing has been considered multifaceted including poor pitch perception, poor motor control, and poor auditory feedback integration (Hutchins & Moreno, 2013). We propose that incorporating measures for performance variability and pitch directionality is a useful approach to seek the origin of poor-pitch singing mechanisms.

8. CONCLUSION

We examined whether singing ability in the pitch-matching task was influenced by the timbre of the stimulus. Both musicians and non-musicians exhibited vocal timbre advantage in comparison to pure tone stimuli. As expected, non-musicians performed more poorly than musicians, in line with previous studies. In the musicians groups, vocalists were not significantly different from instrumentalists, despite their long-term singing experience. This may mean that vocal timbre advantage is universal and robust, but further investigation with better control on subjects' musical experience would be warranted. Our data found that sung pitches were largely flat from the target tones. This may be further investigated by individual adjustment of the target pitch range. Poor-pitch singing in non-musicians was also expressed by the large variability across trials. We propose that this measure is another important indicator of singing ability. Non-musicians also showed larger heterogeneity across individuals. Whether this heterogeneity can be reduced by receiving musical training or not is an important question for future research.

9. REFERENCES

Amir, O., Amir, N., & Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. The Journal of the Acoustical Society of America,113*(2), 1102-1108.*

Boersma, P. (2002). Praat, a system for doing phonetics by computer. Glot international, 5*(9/10), 341-345.*

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2007). Singing proficiency in the general population. The journal of the Acoustical Society of America, 121*(2), 1182-1189.*

Granot, R. Y., Israel-Kolatt, R., Gilboa, A., & Kolatt, T. (2013). Accuracy of pitch matching significantly improved by live voice model. Journal of Voice,27*(3), 390-e13.*

Estis, J. M., Dean-Claytor, A., Moore, R. E., & Rowell, T. L. (2011). Pitch-matching accuracy in trained singers and untrained individuals: the impact of musical interference and noise. Journal of Voice, 25*(2), 173-180.*

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. Journal of Experimental Psychology: General, 141*(1), 76.*

Hutchins, S., & Moreno, S. (2013). The Linked Dual Representation model of vocal perception and production. Frontiers in psychology, 4.

Kostka, S. M., Payne, D., & Schindler, A. (2000). Tonal harmony: with an introduction to twentieth-century music *(Vol. 1221). New York: McGraw-Hill.*

Lévêque, Y., Giovanni, A., & Schön, D. (2012). Pitch-matching in poor singers: human model advantage. Journal of Voice, 26*(3), 293-298.*

Moore, R. E., Estis, J., Gordon-Hickey, S., & Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. Journal of Voice, 22*(4), 399-407.*

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. The Journal of the Acoustical Society of America,125*(1), 328-338.*

Ogawa, Y., & Murao, T. (2004). Flat pitch production by accurate singers: Relationship among pitch discrimination, stimulus model, and pitch-interval matching accuracy. In Proceedings of the 8th International Conference on Music Perception & Cognition *(pp. 723-724).*

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. Journal of neuroscience methods, 162*(1), 8-13.*

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. The Journal of the Acoustical Society of America, 128*(4), 2182-2190.*

Price, H. E. (2000). Interval matching by undergraduate nonmusic majors. Journal of Research in Music Education, 48*(4), 360-372.*

Watts, C., Moore, R., & McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. Journal of Voice, 19*(4), 534-543.*

Watts, C. R., & Hall, M. D. (2008). Timbral influences on vocal pitch-matching accuracy. Logopedics Phonatrics Vocology, 33*(2), 74-82.*

EXPLORATION OF SINGING SKILLS IN PRESCHOOL CHILDREN

Mayumi Adachi[1], Sanae Ogura[1]
[1]Hokkaido University, Japan
Time: 17:30

Understanding singing and the development of singing skills involves examining a multifaceted set of sub-skills (e.g., vocal range, pattern matching, singing a familiar song, generating an original song). In the past, these sub-skills were examined separately. The goals of the present study was to explore relations among singing sub-skills in preschool children, and to investigate how their musical background would influence their singing skills. Twenty-three Japanese children, including 11 boys, 4-7 years of age were tested by means of the AIRS test battery of singing skills (Cohen, et al., 2013). Nine children were taking music lessons. Information on home musical environment was obtained from their mothers. Each testing was audio- and video-recorded. We developed a scoring system for each task, which will be presented at the conference. Due to the small sample size, we used permutation (rather than parametric) t-tests (tobt) for group comparisons. According to our preliminary analyses, children's ability to sing an ascending C-major scale showed relatively high correlations with their abilities to copy short patterns (r_s = .66-.75, ps < .01) and to sing a familiar song with words (r = .69, p < .01) and without words (r = .61, p < .01), but there was no linear correlation with their ability to sing their favorite song (r = .24, p > .10). Children taking music lessons sang both an ascending (M = 7.00, SD = 2.91) and a descending (M = 6.67, SD = 3.94) C-major scales better than their counterpart (M = 3.71, SD = 2.55 for ascending, M = 2.43, SD = 2.59 for descending), tsobt(22) = 2.85.-3.13, ps = .006-.012, d's = .609-.668). The accuracies in singing a familiar song with or without words did not differ as a function of music lessons. However, the proportion of children who sang the familiar song better without words were higher among children with music lessons (67%) than those without music lessons (21%), $\chi 2(1, N = 23) = 4.71$, p = .03. We will report how home musical environment influences preschool children's singing abilities at the conference.

SCHOOL MUSIC PARTICIPATION: EXPLORING THE ROLE OF STUDENTS' SELF-CONCEPT AND SINGING ABILITY

Steven Demorest[1], Peter Pfordresher[2], Jamey Kelley[1]
[1]University of Washington School of Music, USA, [2]University At Buffalo, USA
Time: 18:00

1. Several studies have found that adults who think of themselves as "tone deaf" have normal perceptual abilities, but a perceived lack of ability to sing accurately or well. These adults report that they did not seek further musical training because they felt unmusical. Studies with children have identified musical self-concept as an important factor in students' attitudes toward music and future music participation. It is possible that many children who have difficulty developing the skills of accurate singing may opt out of elective music opportunities. This study examined how children's attitudes about music and about themselves as musicians related to their choice to participate in elective music classes and their actual singing ability.

2. Method: The sample for this study was the entire sixth-grade population of five elementary schools in a single district (n=319). In phase one, surveys were administered to students that explored students' attitudes about music, musical self-concept, and other variables. Survey results were subsequently compared with school registration data, examining students' choices of participation in music elective classes. In phase two, we randomly sampled from the music participant (n=32) and non-participant (n=23) groups and measured their singing accuracy on various singing tasks.

3. Results: 50% of the sample registered for elective music the following year. The initial analysis examined the influence of the following variables: Musical Self Concept, Peer Influence, Perceived Importance of Music, Perceived Cost of Participation, and Enjoyment of Singing. The model predicted students' music participation choices with 70% accuracy. Students who chose to continue in music have significantly more positive responses on all of the variables. There were no significant differences in singing accuracy between the two groups, but accuracy correlated significantly with self-concept across both groups.

4. Discussion: It appears that five variables can predict students' participation choices with 70% accuracy, suggesting that by sixth-grade students are translating their beliefs into action regarding music participation. Given the lack of difference in singing accuracy, the difficulties experienced by some adults may be due to their choices regarding further music training. Those choices may have been driven by self-perceptions of inadequacy not a real lack of skill.

Session 9

[9A] Workshops

SEOUL 09:00- 10:30 Friday 8 Aug 2014

AWAKENING THE KINESTHETIC SENSES IN PIANO PERFORMANCE

Jun Young Kwak[1]
[1]*Korea Piano Pedagogy Association, Korea, Republic of*
Time: 09:00

The basic configuration of the piano has not changed significantly since it was invented in the 18th century. Certainly, important mechanical transformations have taken place. For example, the frame and strings of the piano are more durable because they are made using stronger materials. However, these changes do not mean that the piano has been reborn as a new instrument. Therefore, if one has problems in performing the piano, the major causes of these practical difficulties do not stem from the instrumental design of the piano but rather from misinterpretations of the relationship between the piano and the human body.Many students have had negative experiences with technical studies. These students often attempt etudes without an understanding of exactly what the body can or cannot do. One serious physical problem which every pianist must resolve involves unbalanced finger strength caused by the structure of the hand and its tendons and the dominance of one hand over the other. Etudes cannot "fix" such problems, which are innate to the human body. In addition, students are frequently unable to understand the connection between the etudes and the repertoire they are studying. Even if they devote long periods of practice time to etudes, often no positive results can be observed. This stagnation results in high levels of frustration and feelings of alienation and hopelessness because of the impossibility of rising to the levels of great pianists who seem to have been born with a flawless technique. A fingering does not have inherent meaning. Rather, it becomes meaningful in a musical context. Therefore, teachers should educate themselves about many different possible fingering systems instead of simply relying on personal preference or habitual responses, especially when it comes to assisting their students with fingering.

AN INTERACTIVE INTRODUCTION TO THE MUSIC AND MOVEMENTS OF BRAZILIAN CAPOEIRA

Megha Makam[1], *Blair Kaneshiro*[1], *Jonathan Berger*[1]
[1]*Stanford University, USA*
Time: 09:30

1. Value & Meaning: This workshop propagates the value for which ICMPC is known; it brings together researchers from disparate backgrounds who all share a common passion for understanding our humanness through investigating musical behavior. Ultimately, capoeira is a game, and this workshop will provide conference attendees an informal setting to make music, move together, and have fun embodying a Brazilian spirit.

[9B] Music Therapy: Rhythm and Motor

KYOTO 08:30- 10:30 Friday 8 Aug 2014

MOTOR REHABILITATION FOR STROKE PATIENTS USING MUSICAL RHYTHMIC STIMULATION

Konstantinos Trochidis[1], *Hamzeen Hameem*[1], *Simon Lui*[1]
[1]*Singapore University of Technology and Design, Department of Information Systems Technology and Design, Singapore*
Time: 08:30

1. Background: The majority of patients after a stroke suffer from motor impairments that affect their normal life. Therefore, there is a need for effective rehabilitation of motor functions. Because of this capacity, music has been used for stroke rehabilitation. Despite the increasing evidence in favor of the use of music in stroke rehabilitation, there is a need for better understanding the impact of methods incorporating rhythm and music. This allows the design of more efficient systems for music-supported stroke rehabilitation.

2. Aim: The present study seeks to investigate the rhythmical properties of music on auditory-motor synchronization and enhance the effectiveness of music-supported stroke rehabilitation in a game setting by fully exploiting the emotional and motivational properties of music using physiological feedback.

3. Method: The participants completed a task by tapping synchronous along with the musical beat using a multi-touch music feedback reactable interface based on reacTIVision. The hybrid surface was connected to a projector to provide visual feedback of the tapping movement. The participants completed the task by listening and tapping to different tempi and metrical structures including Classical, Pop and Rock music. During the experiments the tapping movement was detected and three physiological signals, i.e. the electrocardiogram, the respiration rate and the skin conductivity of the listeners were recorded simultaneously to provide insight into the entrainment and the emotional state of the patients.. The phase difference between the tapping movement and inter onset interval of the beats in the music used to assess auditory-motor synchronization. Finally, the Synchrogram method was used to quantify rhythm entrainment between the music stimuli and the cardiorespiratory data.

4. Results: Periods of synchronization were observed between the phase of inter-onset interval of the beat in music and the tapping interval as well as between the tempo of music and the cardiovascular and respiratory measures indicating improved motor function. An effect in skin conductance was observed with an increased level during the musical task compared to baseline indicative of the emotional state of the participants.

5. Conclusions: The findings of this study provide information on how to monitor and improve music-supported therapy programs in a game setting for stroke rehabilitation using music stimulation and physiological indicators to access auditory-motor synchronization. Skin conductivity can be used as a measure to monitor arousal and stress factors during musical training.

MOTOR PERFORMANCE IN POST-STROKE RECOVERY USING ACTIVE MUSIC THERAPY

Päivi Järvinen-Lepistö[1], *Birgitta Burger*[1], *Esa Ala-Ruona*[1]

[1]*Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland*
Time: 09:00

ABSTRACT

Approximately 33 million people worldwide survived a stroke in 2010, though most often being severely disabled afterwards. Effective acute post-stroke rehabilitation is required to treat patients adequately. Music therapy, music listening, and music-based exercises have been shown to improve cognition, mood, and motor functions of stroke patients. This paper suggests that active music therapy could be beneficial for post-stroke recovery and presents a single case investigation that is part of a larger study. After extensive observational analysis of the video material recorded during music therapy sessions, a simple djembe drum pattern was selected to investigate possible improvement in motor performance. Tests at four time points were recorded with an optical motion capture system and computationally analyzed. Results indicate that muscular strength as well as movement activity and control improved in both the left (impaired) and right (unaffected) hand during the recovery process, potentially suggesting a positive effect of active music therapy on post-stroke rehabilitation.

1. INTRODUCTION

The growth and ageing of the global population is leading to a rising number of stroke patients among various age ranges, in particular in ages 20 to 60 in low- to middle-income countries. In 2010, approximately 33 million people survived a stroke (Giroud et al., 2013). Strokes can have various effects including impairment in motor and sensory systems, emotion, language perception, and cognitive functions, causing difficulties in daily life, such as taking care of oneself (Morris & Taub, 2008). The rising amount of strokes requires preventive measures and effective acute post-stroke care, as it otherwise brings high economic costs for societies and is further a leading cause of serious long-term disability of the patient.

More efficient rehabilitation methods have been introduced during recent years in physio- and occupational therapy involving more dynamic, task-oriented, and repetitive training for post-stroke treatment. The idea is that the intensive, meaningful, task-specific training will activate new areas in the cerebral cortex (Zorowitz, 2006; Sivenius & Tarkka, 2008). Such activating forms of therapy based on motor learning principles have been shown to be effective in the recovery of stroke patients' cognitive and physical functions, supporting their return to independent life at home and participation in every-day situations. Stroke rehabilitation involves learning processes related to muscle control and movement skills that enable the patient to live and cope with the continuously changing environment (World Health Organization, 2001, Pyöriä et al., 2007). According to Shumway-Cook and Woollacott (2001), in particular functions of the upper limbs play an important role in post-stroke recovery, being crucial, for instance, in walking or in keeping balance.

The use of music and music therapy in clinical contexts was found to positively affect cognitive, emotional, and motor abilities of impaired clients (Schlaug et al., 2010). With help of brain measure techniques, such as electroencephalography (EEG) or magnetoencephalography (MEG), it was shown that music listening as well as making activates a complex network of brain areas related to auditory, cognitive, sensory-motor, and emotional processes (e.g., Ellis & Thayer, 2010; Särkämö et al., 2013), In particular related to stroke rehabilitation, it has been demonstrated that playing musical instruments can effectively improve motor skill recovery (Altenmüller et al., 2009; Schneider et al., 2007, 2010; Thaut & Abiru, 2010).

Schneider et al. (2010) conducted a study with stroke patients and compared music-supported training (playing on an electric piano or drums) with constraint-induced movement therapy (CIMT – a therapeutic approach focusing on restraining the unaffected limb while intensively using the affected limb, see Morris & Taub, 2008) and conventional physiotherapy. By analyzing motor functionality with both a set of established motor tests and a computer-based movement analysis system, they found that music-supported training was more effective in terms of recovery of fine motor functions compared to CIMT and conventional physiotherapy.

Additionally, patients described music-supported training as enjoyable and motivating, which has been identified as playing an important role in a successful recovery process (Schneider et al. 2010). Furthermore, Koelsch et al. (2010), as well as Forsblom et al. (2010) and Särkämö et al. (2008) have found positive effects of making and listening to music on mood and motivation in their studies with people with affective disorders and stroke patients.

2. AIM

The aim of this study was to investigate if active music therapy (i.e., the client being actively involved in musical interaction with a therapist, see below), added to standard care, has a positive impact on the recovery of motor functions after stroke. With the objective of evaluating this new model of active music therapy for stroke rehabilitation, we started from observing and finding characteristic features in the patient's therapy process. This procedure enabled us to identify effects and detect possible changes in the motor recovery process, and to subsequently analyze them in measurable ways. While being part of a larger study, this article will present initial results from a single case investigation.

3. METHOD

In the following, the research design, the clinical intervention model, and the specific methodological aspects regarding the case study approach will be explained.

3.1. Research design and clinical intervention

Forty-five patients suffering from right hemisphere middle cerebral artery stroke (diagnosis ensured with magnetic resonance imaging – MRI) participate in the study (Ala-Ruona, 2009). They receive standard care and additionally two weekly sessions of individual active music therapy with a clinically trained and further educated music therapist over a period of three months (20 sessions) (Forsblom & Ala-Ruona, 2012). Clients are randomly assigned into receiving music therapy as an early intervention during months 1 to 3 or during months 4 to 7 after stroke as a delayed intervention (cross-over RCT). Standard care, following the Finnish Current Care Guidelines (see http://www.kaypahoito.fi/web/english) for stroke, is given during the whole period of time. The active music therapy model is a combination of structured and non-structured musical exercises and includes rhythmic tasks of different levels of difficulty, music-assisted relaxation, interactive improvisation and therapeutic discussion. Test sessions at four time points (baseline, after three, four, and seven months) are recorded using an optical motion capture system.

3.2. Specific aspects regarding the case study

Participant: The patient (male, age: 57) was diagnosed with an acute right hemisphere middle cerebral artery stroke and assigned into active music therapy treatment during months 1 to 3 after stroke. The baseline evaluation tests in the hospital showed that he was not able to walk without help, having furthermore severe problems in using the upper extremities in self-care like dressing and washing himself. We used the video-recorded baseline test session to conduct an observational analysis of the client's functioning. The client was able to independently walk small distances in the test room. He had large difficulties in using the left upper extremity when playing drums. The ability to selectively activate distal muscles in his affected arm was limited. Furthermore, the strength was found reduced, the range of wrist motion was severely limited, and the finger opening for objects was insufficiently. He was unable to grasp or release objects with the left hand. He supported lifting and reaching movements of the left upper extremity with the right hand or increased the trunk movement to swing the arm. Occasionally, he used the right hand to lift and drop the left hand onto the djembe drum to forcefully achieve opening of the fingers and relaxation of the hand. Tension was found when trying to move hand and finger muscles passively. During the test session, the left upper extremity showed increasing fatigue, with movements getting smaller, slower, and more uncontrolled.

Task: Various motor performance tests consisting of rhythmic patterns were played on djembe drums together with a test therapist, as well as on a special drum set with the test therapist playing piano. The level of challenge in the tasks was adjusted according to the client's level of functioning. For the current analysis of this single case, a simple 3-hits + pause (I I I _ I I I _ …) pattern – first performed several times with the right hand, then with the left hand – was selected to be able to investigate possible improvement in motor performance of both left and right hands.

Procedure of the analysis: In order to understand the recovery process, and to recognize meaningful movement patterns indicating measurable changes in motor performance, an extensive video analysis of all the 20 therapy sessions was conducted. After thorough analysis, we decided to focus on one of the simplest rhythmic pattern in the test battery (the three-hit pattern mentioned above), and to analyze the overall movement range, wrist and finger movements, dorsiflexion of the wrists, and kinetic energy of upper extremity of both hands.

Apparatus: The client's movements were recorded using an eight-camera optical motion capture system (Qualisys ProReflex), tracking, at a frame rate of 120 Hz, the three-dimensional positions of 19 reflective markers attached to the client's body. The locations are depicted in Figure 1A. Additionally, four Sony video cameras were used to record the test sessions enabling the observational analysis. The audio data from the djembe drums was recorded into ProTools software.

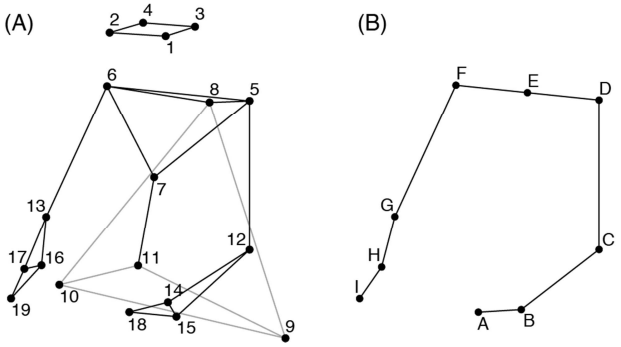


Figure 1. (A) Client's marker setup; (B) Joints and segments used in the analysis

Movement data processing: The MATLAB Motion Capture (MoCap) Toolbox (Burger & Toiviainen, 2013) was used to process and analyze the movement data. First, the data was trimmed to a 10-second excerpt of the task of interest after removing the first sequence of the 3-hits pattern to ensure a stable execution of the pattern. Next, a set of nine secondary markers – subsequently referred to as joints – was derived from the original set of markers. This was done to reduce the amount of markers, to exclude unnecessary ones, and to create the chain of joints required for body segment modeling (see below). The locations of the nine joints are depicted in Figure 1B. For seven joints (A, C-G, and I), the locations were identical to one of the original markers, while the wrist joints (B and H) were obtained by averaging the locations of markers 14 and 15 as well as 16 and 17. This chain of joints was subsequently used to model body segments (i.e., the connection between two joints) based on Dempster's parameters (Dempster, 1955) to calculate angles between segments and kinetic energy contained in the body segments (Toiviainen et al. 2010).

4. RESULTS

This analysis comprises comparisons between Time Points (TPs) 1, 2, and 4. TP 3 was excluded, since we aimed at focusing on the immediate effects of the active music therapy and the long-term sustainability of the effects. In order to investigate hand dorsiflexion, the vertical displacement of wrist and finger joints of left and right hand (joints A, B, H, and I) was compared and plotted as a function of time (see Figure 2). The 3-hits pattern is visible in both hands and all TPs; however, in TP 1, both left and right hand exhibited the smallest movement ranges, while they were highest in TP 2 (see Table 1 for standard deviations of finger displacement). The displacement range of both hands was similar in each TP (see SDs in Table 1). In TP 1, wrist and finger time series specifically of the left hand were almost identical (i.e., the hand being mainly parallel to the floor), while they were more dissimilar in the other TPs. Furthermore, the left hand movement peaks in TP 1 have rather variable heights, whereas they became more regular in TPs 2 and 4. The same applies to the right hand, albeit the displacement pattern was more regular in general, changing from a somewhat arbitrary peak structure in TP 1 to a strong-weak-weak pattern in TP 2 to a strong-weak-strong pattern in TP 4. The displacement curves of both hands in TP 1 show resting periods after the drumming pattern was completed. This was, however, due to the first instruction the client received on how to play the pattern, as it is supposedly easier to keep the hand resting on the drum to stay in time instead of bouncing back.

Table 3. Standard deviations of vertical finger displacements

	TP 1		TP 2		TP 4	
	right	**left**	**right**	**left**	**right**	**left**
SD	40.92	35.34	88.24	78.67	68.02	56.10

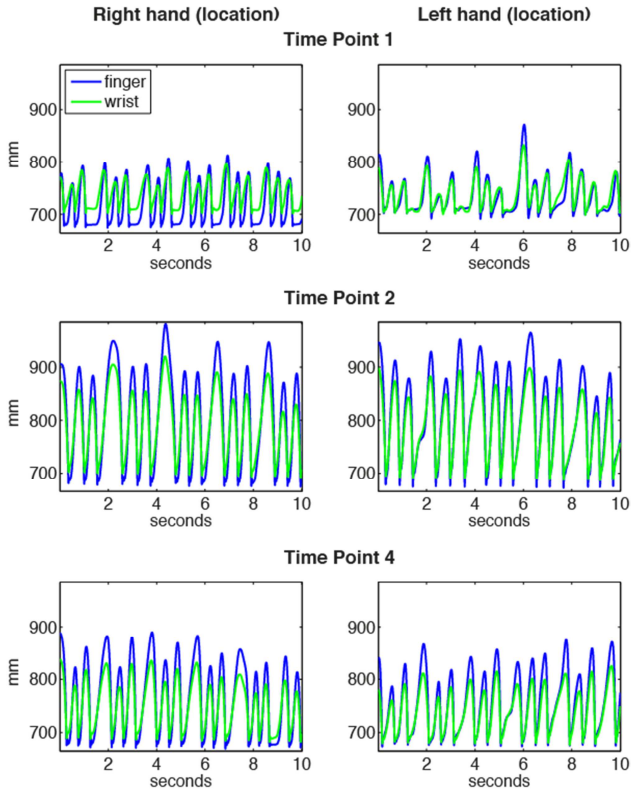


Figure 2. Vertical displacement of wrist (green) and finger (blue) of right and left hand for TPs 1, 2, and 4.

In order to further investigate dorsiflexion, hand angles (segments A-B and H-I) were calculated (see Figure 3). While the averages of the angle time series were found to be relatively similar (between 18.06° and 22.78°), the shapes of the curves differed notably. The curves for the left hand appear more irregular, jerky, and unstable than the right hand ones. Only in TPs 2 and 4, the 3-hits pattern is recognizable, whereas for the right hand, the 3-hit pattern is visible in all three TPs. In TP 1, the left hand's angle decreased with time; in other words, the client was less flexing the left hand the longer he played. Right hand angles differ in particular in range, with TP 1 having the smallest range (18.47°, SD: 4.43), followed by TP 4 (21.98°, SD: 8.86), and TP 2 having the largest range (36.75°, SD: 5.31).

Kinetic energy of the hands (segments A-B and H-I) is displayed in Figure 4. It was found to be least in TP 1 and highest in TP 2, in particular for the left hand. The 3-hits pattern can be seen in all TPs and both hands (least clear for TP 1 left hand). In TPs 2 and 4, the peaks of the right hand imply clear accent patterns – however, different ones for each TP: TP 2 shows an increasing use of energy per hit (within the three-hit pattern), whereas in TP 4, the energy contained in the third hit was strongest, followed by the first hit and the weakest being the second hit. The peak structure of the left hand for TP 2 and 4 were more irregular in that respect. Furthermore, the right hand in TP 2 and 4 as well as the left hand in TP 4 were similar in range.

5. DISCUSSION

The presented study investigated motor impairment and its recovery processes during rehabilitation after stroke using a combination of standard care and active music therapy. This article in particular examines the use of hands and the relation between wrist and finger in a simple djembe-drumming task. Using optical motion capture data, we analyzed vertical displacement of finger and wrist of both left and right hand, angles of both hands, and the kinetic energy contained in the hands at three distinct time points within the first seven months after stroke onset. Overall results of these analyses (vertical displacement, angles range, and kinetic energy) indicate a general pattern of lowest movement activity in Time Point 1, highest activity in Time Point 2, and medium activity in Time Point 4. These results suggest a major impairment of motor functions after stroke, an over-activation of (muscular) power and energy during the early recovery process – maybe related to retrieving the muscular strength, but not being able to control it – and finally, after seven months, being able to control and more deliberately use the recovered power. Over-activation of upper limbs in an early stage after stroke has been reported in previous literature (Corbetta, 2008), thus it seems to be a common phenomenon. It might be possible to link the regained control after seven months to a sustainable long-term effect. When thinking about transfer effects between therapy and everyday life, the objectives, goals, and rehabilitative tasks done in the therapy should be in line with the skills needed in daily living in the patient's own environment. The patient should be able to utilize the skills learned during the therapy, at best being able to achieve sustainable long-term effects. Seven months after the stroke, the patient only showed slight difficulties in daily activities. However, more research is needed to find how the transfer effect from the improvement in the therapeutic process can be integrated into participation in everyday activities and the overall quality of life.

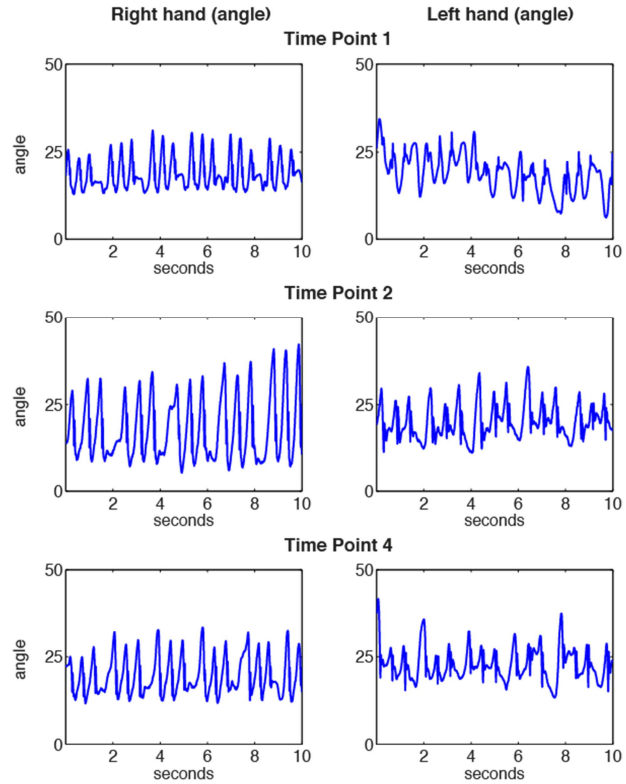


Figure 3. Angles of right and left hand segments at TP 1, 2, and 4.

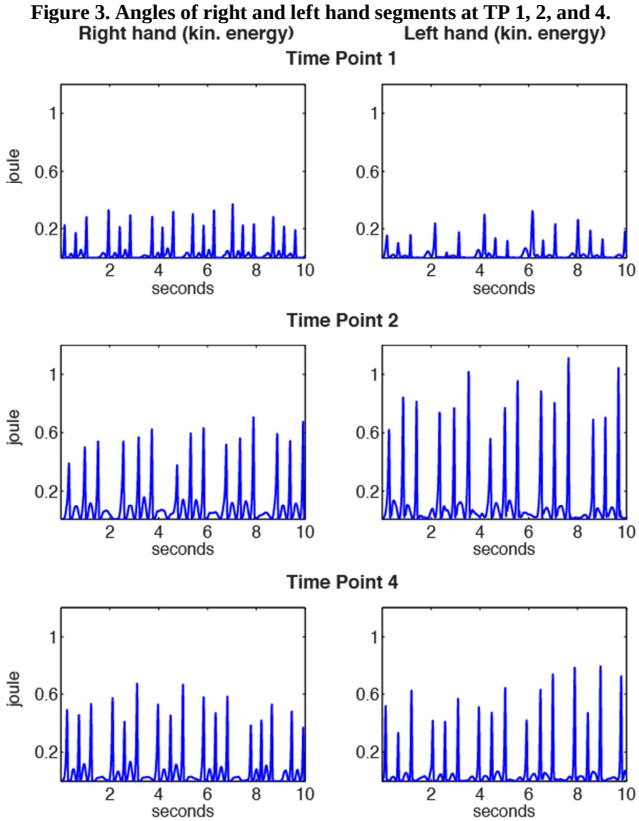


Figure 4. Kinetic energy contained in left and right hand segment.

Regaining movement control is one of the crucial components in post-stroke recovery. The variable heights of peaks in the vertical displacement for the left hand in all time points and for the right hand in Time Point 1 imply that the client was unable to control his movement in the beginning, but gained more control over time, indicated by the increasing regularity of the peaks in the later time points. The uncontrollability

of the left hand is visible particularly in the angles, as the angle time series are considerably noisier than the right hand ones. The disparity related to the angles could be used as an indicator to distinguish between left and right hand.

The almost identical vertical displacement of both hands found in Time Point 1 suggests that the client failed to perform hand dorsiflexion, indicating that his hands and arms were very stiff and inflexible after the stroke, while becoming more flexible during the recovery process making dorsiflexion possible. Furthermore, the decrease of the left hand angle indicates that the client was unable to flex the hand but kept it instead rather straight, the longer the task was proceeding. This could suggest a fatigue effect.

The results of our study suggest that the bilateral use of both hands improves post-stroke recovery. McCombe Waller and Whitall (2004) have found that bimanual training improved inter-limb coordination of the affected limb as well as control of the unaffected side. Furthermore, Stewart et al. (2006) reported bilateral movement training to be beneficial either alone or in combination with rhythmic cues. Besides elements such as the djembe task analyzed here, in which the patient plays patterns using one hand only, the active music therapy model also includes several other tasks (performed on a drum kit), in which the patient plays specific patterns with both hands simultaneously. Investigating such tasks will provide further insights into possible positive effects of bilateral movement training.

Motivation and feedback are both essential features in motor learning. We believe that active music therapy is a very useful way of getting the patient engaged in goal-oriented and meaningful tasks, as they motivate the patient to continue, maybe even going beyond boundaries. Interaction and interactive play with the therapist are perceived as supportive and encouraging (Forsblom & Ala-Ruona, 2012). It is further assumed that, as long as the patient is actively involved in such tasks, neuromuscular processes – being essential for motor learning and recovery – should get enhanced.

6. REFERENCES

Ala-Ruona, E. (2009). Active music therapy for post-stroke recovery. In Current Clinical Trials (ISRCTN03493594). (http://www.controlled-trials.com/ISRCTN03493594)

Altenmüller, E., Marco-Pallares, J., Münte T. F., & Schneider, S. (2009). Neural reorganization underlies improvement in stroke-induced motor dysfunction by music-supported therapy. Annals of the New York Academy of Sciences, 1169, 395-405.

Burger, B., & Toiviainen, P. (2013). MoCap Toolbox – A Matlab toolbox for computational analysis of movement data, in Proceedings of the 10th Sound and Music Computing Conference, Stockholm, Sweden, 172-178.

Corbetta, M. (2008). Functional brain imaging and neurological recovery. In D. T. Stuss, G. Winocur, & I. H. Robertson (Eds.), Cognitive neurorehabitation. Evidence and application (pp. 163-175). New York: Cambridge University Press.

Dempster, W. T. (1955). Space requirements of the seated operator: Geometrical, kinematic, and mechanical aspects of the body with special reference to the limbs. WADC Technical Report 55-159, Wright-Patterson Air Force Base, Ohio.

Ellis, R. J. & Thayer, J. F. (2010). Music and autonomic nervous system (dys)function. Music Perception, 27(4), 317-326.

Forsblom, A., & Ala-Ruona, E. (2012). Professional Competences of Music Therapists Working in Post-stroke Rehabilitation. Voices: A World Forum For Music Therapy, 12(3). (https://normt.uib.no/index.php/voices/article/view/ 647/561)

Forsblom A., Särkämö, T., Laitinen, S., & Tervaniemi, M. (2010). The effect of music and audiobook listening on people recovering from stroke: The patient's point of view. Music and Medicine, 2(4), 229-234.

Giroud, M., Jacquin, A., & Béjot, Y. (2014). The worldwide landscape of stroke in the 21th century. The Lancet, 383(9913), 195-197.

Koelsch, S., Offermanns, K., & Franzke, P. (2010). Music in the treatment of affective disorders: An exploratory investigation of a new method for music-therapeutic research. Music Perception, 27(4), 307-316.

McCombe Waller, S., & Whitall, J. (2008). Bilateral arm training: Why and who benefits? NeuroRehabilitation, 23(1), 29-41.

Morris, D., M., & Taub, E. (2008). The use of constraint- induced movement therapy (CI therapy) to promote motor recovery following stroke. In D. T. Stuss, G. Winocur, & I. H. Robertson (Eds.), Cognitive neurorehabilitation. Evidence and application (pp. 401-413). New York: Cambridge University Press.

Pyöriä, O. (2007). Reliable clinical assessment of stroke patients' postural control and development of physiotherapy in stroke rehabilitation. PhD Thesis. Studies in Sport, Physical Education and Health 125, Jyväskylä: University of Jyväskylä.

Särkämö, T., Tervaniemi, M., & Huotilainen, M. (2013). Music perception and cognition: Development, neural basis, and rehabilitative use of music. WIREs Cognitive Science, 4, 441-451.

Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., Autti, T., Silvennoinen, H. M., Erkkilä, J., Laine, M., Peretz, I., & Hietanen, M. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. Brain, 131(3), 866-876.

Schlaug, G., Altenmüller, E., & Thaut, M. (2010). Music listening and music making in the treatment of neurological disorders and impairments. Music Perception, 27(4), 249-250.

Schneider, S., Münte T., Rodriquez-Fornells, A., Sailer, M., & Altenmüller, E. (2010). Music-supported training is more efficient than functional motor training for recovery of fine motor skills in stroke patients. Music Perception, 27(4), 271-280.

Schneider, S., Schönle, P.W., Altenmüller, E., & Münte, T. (2007). Using musical instruments to improve motor skill recovery following a stroke. Journal of Neurology, 254(10), 1339-1346.

Shumway-Cook, A., & Woollacott, M. H. (2001). Motor Control-Theory and Practical Applications. Baltimore: Williams & Wilkins.

Sivenius, J., & Tarkka, I. M. (2008). Aivovaurioiden kuntoutuksen neurobiologinen tausta [Neurobiological background of rehabilitation for brain damage]. In P. Rissanen, T. Kallanranta, & A. Suikkanen (Eds.), Kuntoutus [Rehabilitation] (pp. 289-297). Keuruu: Duodecim.

Stewart, K. C, Cauraugh, J. H., & Summers, J. J. (2006). Bilateral movement training and stroke rehabilitation: A systematic review and meta-analysis. Journal of the Neurological Sciences, 244(1-2), 89-95.

Thaut, M. H., & Abiru, M. (2010). Rhythmic auditory stimulation in rehabilitation of movement disorders: A review of current research. Music Perception, 27(4), 263-269.

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. Music Perception, 28(1), 59-70.

World Health Organization (2001). The International Classification of Functioning, Disability and Health (ICF). Geneva: WHO. (http://www.who.int/classifications/icf/en)

Zorowitz, R. D. (2006). Neurorehabilitation of the stroke survivor. In M. Selzer, S. Clarke, L. Cohen, P. Duncan & F. Gage (Eds.), Textbook of neural repair and rehabilitation: Medical neurorehabilitation (pp. 579-592). Cambridge UK, Cambridge University Press.

ROBOT-ASSISTED ARM TRAINING AND RHYTHMIC ACOUSTIC STIMULATION: THE ROLE OF DIFFERENT RHYTHMIC STIMULI ON MOTIVATION AND PERFORMANCE IN MOTOR-REHABILITATION TRAINING AFTER STROKE

Florina Speth[1], Uwe Seifert[2], Michael Wahl[1]
[1]Humboldt University Berlin; Institute of Rehabilitation Science, Germany, [2]University Cologne; Institute of Musicology, Germany
Time: 09:30

ABSTRACT

Robotic arm rehabilitation training (RAT) intensifies and complements conventional ergo therapy for post-stroke arm paresis treatment effectively. Usually robotic devices are connected to multimedia scenarios including sound. In arm training utilizing sound, it was shown that patients benefit from Rhythmic Acoustic Stimulation (RAS). Up to now, neither effects of sound within RAT were examined, nor a specification of RAS for RAT was developed or investigated. Three pilot experiments (E1–3) were conducted with 20 healthy subjects performing the Nine Hole Peg Test throughout specified rhythmic stimulation designs (waltz-music, metronome-, multisensorial-, spearcon-beat), and a condition without stimulation to evaluate effects on function and motivation. E1–3 differed in degree of difficulty to simulate limitations caused by arm paresis. Time, performance, and mood were ranked, and the two strongest designs were determined and further discussed for the application to RAT. In all experiments time and performance were better with rhythm than without. In E1 best results were achieved with waltz-music, in E2 and E3 performance and time were best with metronome-beat. Mood was rated highest for waltz-music throughout all experiments. Generally, our results suggest that RAS enhances performance and time. To promote function and motivation in RAT a combination of waltz-music with metronome-beat is suggested for further investigations applied in RAT with patients.

1. INTRODUCTION

Currently there is an increasing research interest in neurology and technology on the effectiveness of music applied in motor rehabilitation training for stroke patients (Altenmüller & Schlaug, 2013; Rosati et al., 2011). 90% of all stroke patients suffer from a loss of arm function. As in 30–40% of these cases the syndrome resists chronic, there is a need for innovative therapeutic strategies counteracting a life-long disability. Studies on effects of different therapeutic techniques showed that an early start of highly intensive active training in which goal-directed tasks are practiced repetitively, gain best results ((Kwakkel, Kollen & Krebs, 2008; Platz 2003). To enrich conventional therapy and to increase training frequency with high rates of movement-repetition, robot-assisted arm training (RAT) is regarded as very promising, especially for the heavy exposures of the paresis (Hesse et al. 2004; Prange et al. 2006). In RAT the robotic device usually serves as controller within a virtual training environment displaying performance qualities and offering playful training with games. As one part of this multimedia scenario, sound appears within interactive video-game sounds, background music and performance feedback (Rosati et al. 2011). Till now to the best of our knowledge the role of sound and music within RAT was never investigated, or specified by taking previous research on effects of music in post-stroke motor rehabilitation into account. Generally, music and sound were already utilized in a range of motor rehabilitation techniques showing strong effects in motivational as well as functional aspects, especially with the therapeutic technique “Rhythmic Acoustic Stimulation” (RAS) and with music (Bradt et al., 2010; Malcom et al., 2009). This paper presents a review on effects of sound and music in post-stroke motor rehabilitation .In addition, promising effects are suggested for the application in RAT. Our hypothesis is that RAS and rhythmical music applied in RAT might enhance function and motivation. As RAS was never specified for this purpose, different rhythmic stimulation designs were developed and their effects on function and motivation during performance of a fine-motor task were explored in three pilot-experiments. The goal was to determine two promising rhythmic stimulation designs under four different designs in comparison to no stimulation. Further on results of the experiments are discussed for application to RAT. The overall aim of this study is to provide rhythmic stimulation designs for post-stroke RAT improving recovery.

2. MUSIC IN POST-STROKE MOTOR REHABILITATION TRAINING

Music and sound are used in a range of therapeutic techniques in rehabilitation training for stroke patients suffering from hemiparesis (Altenmüller & Schlaug, 2013): Most significant results of music and sound were shown in gait- and arm-training utilizing rhythmical music like march music or a metronome with a therapeutic technique called “Rhythmic Acoustic Stimulation” (RAS) (Bradt et al. 2004). Patients benefit from rhythm-assisted motor training because rhythmic cues serve as external time-keeper supporting movement initiation and synchronization. Studies on RAS-gait training showed that significant improvements in motor qualities were reached in comparison to standard physiotherapeutic treatments (Malcom et al. 2002; Thaut et al. 2002). A pilot study examined effects of bilateral arm-training with rhythmic auditory cueing (BATRAC). After six weeks BATRAC-training motor performance, isometric strength and range of motion improved (Whitall et al. 2009). Another pilot-study investigated effects of RAS applied in arm reaching training. Results showed a significant decrease in compensatory trunk movement, an increase in shoulder flexion, and a slight increase in elbow extension. Movement timing and velocity improved significantly (Malcom et al. 2002). Schneider et al. (2011) observed effects of active music practice with a technique called “Music supported Therapy” (MST) on 20 chronic stroke patients suffering from arm paresis. After 15 sessions applied additionally to conventional treatment, patients showed significant improvements in speed, precision, movement smoothness and activities of daily living. Friedman et al. (2014) carried out a study to evaluate functional and motivational effects on chronic stroke patients with moderate hand paresis of hand function training with the technical interface “Music-Glove” in comparison to conventional hand therapy, and isometric tasks, in a within-subjects design. Results showed that Music-Glove training lead to stronger improvements in hand function and it gained higher motivation ratings than conventional therapy and isometric task training. Results sustained one month follow-up (Friedman et al. 2014). Following, the application of RAS and music applied in RAT might be a promising starting point as it potentially promotes function. Music may raise motivation during training. Our hypothesis is that RAS and rhythmical music applied in RAT enhances function and motivation.

3. METHODS

Goals: To investigate if RAS applied to RAT influences function and motivation, three pilot-experiments (E1-3) were performed. The first goal was to investigate effects of four different rhythmic stimulation designs varying in signal complexity, meter, sound material and modality compared to no stimulation on healthy subjects applied during performance of a fine-motor task. Difficulty raised from an easy level E1, to more difficult levels E2 and E3. In E2 and E3 technical limitations were provided to challenge healthy subjects like patients are demanded due to arm

paresis. The second aim of E1-3 was to identify designs achieving strongest effects to suggest specified RAS-designs for further investigations with patients.

Participants: 20 subjects took part in E1-3 (female=11; male=9; mean age=27.15). 19 participants had previous experiences in playing a musical instrument. All participants were right handed.

Technical set-up: In E1-3 the Nine Hole Peg Test (NHPT), a validated clinical assessment tool for specification of finger dexterity (tip pinch), was performed. In this test, duration and performance to completion of the task reflect manipulative fine-motor skills, especially the ability to execute the pinch grip (Mathiowitz et al. 1985). To assess motivation, a self-evaluation questionnaire was performed, in which participants rated their mood via Visual-Analog-Scale (VAS) (-10; +10) in relation to their initial mood (condition 0) after every single task. In E1 normal test-conditions were given, whereby in E2 and E3 technical limitations were applied: In E2 the test was performed against force of elastic ropes pulling backwards. By this the extensor muscles of arm and fingers had to work harder than in a normal condition. Due to arm paresis most patients suffer from spasticity decreasing force control of extensor muscles (Thibaut et al. 2013). In E3 a mechanical grasp arm was used to perform the task. Like this a loss of precision and coordination given in patients due to arm paresis was simulated. In RAS a metronome or march-music with a 2/2 or 4/4 time signature is commonly used providing intuitive beat synchronization for bipedal gait. As articulation in unilateral arm movements differs from walking, alternatives to commonly used RAS designs were provided: To explore another meter, waltz-music with a triple meter was chosen. Waltz-music might be associated with social interaction during partner dance in which gestural communication via hands is central. Dingler et al. (2011) found that spearcons were learned faster than more abstract earcons applied in auditory displays. To present a rhythmic stimulation design contrasting to abstract sound cues like metronome and a complex musical stimulus like waltz-music, spearcons served as material for another design utilizing speech sounds. A multisensorial rhythm was applied to investigate if effects differ from purely auditory cues. A tempo of 200 bpm was chosen for all designs to provide a speed up rate of 20% in relation to NHPT-time standard value table. The stimulation- conditions were a) no stimulation, b) metronome, c) spearcon-beat: processed audio samples of words like “Super”, “Great”, d) waltz-music: “Voices of spring”, J. Strauss, e) multisensorial-beat: rhythmical hits on the foot and metronome.

Procedure: Before the experiments started, participants rated their initial mood (condition 0). All participants were informed about the test procedure with a pre-recorded audio guide. Before each experiment participants were introduced to the task and given condition by audio guide again. For each test condition (E1-3) one test trial was performed before data recordings started. During the test performance of E1-3, four stimulation designs (condition b-e) were applied and compared to a condition without stimulation (condition a): The stimulation designs a-e were displayed in each test condition in randomized order. Duration till task completion was measured with a stopwatch up from first to last peg contact. Performance quality was assessed counted amount of mistakes. After each trial participants rated their mood in relation to their initial mood. A paired t-test (level of significance: 0.1%) was performed,

4. RESULTS

Experiment 1: Duration was best with waltz-music (mean time (condition d) = 16.6s), followed by spearcon-beat (mean time (condition c) = 16.8s) and weakest without stimulation (mean time (condition a) = 18.3s). (SD = 0.95). Effects were not significant. Performance quality was best with waltz-music (mean performance (condition d) = 0.1) and with multisensorial-beat (mean performance (condition d) = 0.1) and weakest without stimulation (mean performance (condition a) = 0.35). (SD = 0.27). Effects were not significant. Mood was rated best during waltz-music (mean mood (condition d) = 3.67) followed by no stimulation (mean mood (condition a) = 2.57). In relation to the preassessed initial mood (mean mood (0) = 3.65), waltz-music gained still higher values. Weakest results were seen with multisensorial-beat (mean mood (condition e) = -0.04). (SD = 1.37). The effect of waltz-music on mood ratings lead to significantly better mood ratings than multisensorial-beat (p-value = 0.392).

Experiment 2: Best duration was measured with metronome (mean time (condition b) = 25.1s) and weakest without stimulation (mean time (condition a) = 32.1s). (SD = 0.77). Effects were not significant. Performance qualities were best with metronome (mean performance (condition b) = 0.1) and with multisensorial-beat (mean performance (condition d) = 0.1) and weakest without stimulation (mean performance (condition a) = 0.4). (SD = 0.26). Effects were not significant. Mood was best during waltz-music (mean mood (condition d) = 2.14) whereby the initial mood was still higher (mean mood (condition 0) = 3.65). Weakest results were gained with spearcon-beat (mean mood (condition c) = -0.59). (SD = 1.32). Effects were not significant. Experiment 3: Best duration was assessed with metronome (mean time (condition b) = 73.5s), followed by spearcon-beat (mean time (condition b) = 73.6s). Weakest duration was measured without stimulation (mean time (condition a) = 105.6s). (SD = 12.12). Effects were not significant. Performance qualities were best with metronome (mean performance (condition b) = 1.05) followed by multisensorial-beat (mean performance (condition d) = 1.15) and weakest without stimulation (mean performance (condition a) = 1.4), followed by waltz-music (mean performance (condition d) = 1.35). (SD = 0.13). Effects were not significant. Mood was rated highest with waltz-music (mean mood (condition d) = 2.09), whereby the initial mood was still higher (mean mood (condition 0) = 3.65). Weakest results were assessed with spearcon-beat (mean mood (condition c) = -0.25). (SD = 1.35). Effects were not significant.

5. DISCUSSION

In E1-3 duration and performance were weakest without stimulation in comparison to additional stimulation. This may indicate that RAS enhances velocity and performance qualities if applied during a fine-motor task in healthy subjects, independently of the level of difficulty. Mood was rated best with waltz-music in comparison to no stimulation, whereby the initial mood, before any task was performed, was still better in E2 and E3. In E1 mood was rated higher with waltz-music than the initial mood. Waltz-music potentially enhanced the mood whether a task had to be performed actively. This suggests that music was perceived as motivating generally. Mood was rated second best with metronome. In easy conditions given E1, the second best results in mood ratings were seen without stimulation. In more difficult tasks (E2, E3), mood was second best with metronome. Potentially the metronome was perceived as most neutral during demanding tasks. This could be related to results gained with metronome in E2 and E3 in which metronome gained best results in duration and performance. To provide sound designs for RAT enhancing recovery, function and motivation have to be taken into account carefully: Multisensorial-beat led to lower qualities in comparison to other designs in all conditions. This could be due to technical limitations of the robotic prototype generating hits on the foot that were described as distracting by many participants. As mood ratings were strongly negative under this condition, and effects on function were weaker than other designs, it was excluded for further investigations. No strong effects were seen with spearcon-beat, which was described as distracting due to aesthetical reasons. Because of that it was excluded. A simple stimulus like a metronome could serve as a stable rhythmic cue during difficult tasks if provided in an adequate tempo. In all experiments mood was rated highest with waltz-music independently of the level of difficulty. As a

special focus in research on rehabilitation robotics is the proposition of an engaging technical environment that promotes motivation, effects of waltz-music on mood seen in the experiments should be considered as very important.

6. CONCLUSION

The goal of this study is to provide rhythmic stimulation designs promoting functional and motivational effects applied in RAT for stroke patients. A review on effectiveness of sound and music in motor rehabilitation training was taken into account to translate promising stimuli into the context of RAT: As Rhythmic Acoustic Stimulation (RAS) was shown to be highly efficient in enhancing motor control and movement initiation for neurological patients and musical stimulation showed good results in influencing motivation and function during motor training, three pilot-experiments were carried out to examine effects of specified rhythmic stimulation designs and rhythmical music for RAT. The rhythmic stimulation designs consisted of rhythmical music, metronome, a speech-based rhythmical pattern and a multisensorial beat. Three experiments were performed varying in level of difficulty. Throughout three experiments with healthy subjects performing a fine-motor task under four stimulation designs and a base-line condition without stimulation, results showed that specified RAS-designs enhanced velocity profiles and performance in all suggested designs in comparison to no stimulation. Metronome stimuli lead to better performance, especially throughout difficult tasks. In the easiest task, best functional results were gained with waltz-music. Stimulation with waltz-music resulted in best mood ratings throughout each experiment, independently of the degree of difficulty. Function was weaker with waltz-music in more difficult conditions than with metronome or the spearcon-beat. The next goal is to evaluate effects of metronome, and a combination of waltz music with metronome applied during a fine motor test with stroke patients, and further on during robot-assisted hand function training. Further research on effects of sound and music applied in technology-assisted motor rehabilitation is needed to offer therapeutic meaningful auditory stimulation within a complex therapeutic environment that aims to engage motivation during training and to promote recovery effectively.

7. REFERENCES

Altenmüller, E. & Schlaug, G. (2013). Neurologic music therapy: The beneficial effects of music making in neurohabilitation. Acoustical Science & Technology, 34; 1, 5–12.

Bradt, J., Magee, W.L., Dileo, C., Wheeler, B.L. & Mc Gilloway,E. (2004). Music therapy for acquired brain injury (Review). The Cochrane Collaboration, *1–42.*

Dingler, T., Lindsay, J. & Walker, B.N. (2008). Learnability of sound cues for environmental features: auditory icons, earcons, spearcons, and speech. Proceedings 14th International Conference on Auditory Display, *2008, 1-6.*

Friedman, N., Chan, V., Reinkensmeyer, A.N., Beroukhim, A., Zambrano, G.J., Bachmann, M. & Reinkensmeyer, D.J. (2014). Retraining and assessing hand movement after stroke using the MusicGlove: comparison with conventional hand therapy and isometric grip training. Journal of NeuroEngineering and Rehabilitation, *11; 76, 1-14.*

Hesse, S., Werner, C. & Bardeleben, A. (2004). Der schwer betroffene Arm ohne distale Willküraktivität – ein „Sorgenkind" der Rehabilitation nach Schlaganfall!?!. Neurologie & Rehabilitation, *3, 120-126.*

Kwakkel, G., Kollen, B.J. & Krebs, H.I. (2008). Effects of Robot-Assisted Therapy on Upper Limb Recovery After Stroke: A Systematic Review. Neurorehabilitation and Neural Repair, 22: 2, 111–121.

Malcom, M.P., Massie, C. & Thaut, M.H. (2002). Rhythmic Auditory-Motor Entrainment Improves Hemiparetic Arm Kinematics During Reaching Movements: A Pilot Study. Topics in Stroke Rehabilitation, *16, 69–79.*

Mathiowitz, V., Weber, K., Kashman, N. & Volland, G. (1985). Adult Norms for the Nine Hole Peg Test of Finger Dexterity. The Occupational Therapy Journal of Research, *5; 1, 24-38.*

Platz, T.H. (2003). Research: Evidence-based arm rehabilitation – systematic review of the literature. Der Nervenarzt, *74; 10, 841–849.*

Prange, G.B., Jannink, M.J.A. & Groothuis- Oudshoorn, C.G.M. (2006). Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke. Journal of Rehabilitation Research & Development, *43; 2, 171-184.*

Rosati, G., Reinkensmeyer, D.J., Avanzini, F., & Masiero, S. (2001). Improving robotics for neurorehabilitation: enhancing engagement, performance, and learning with auditory feedback. Proceedings of the IEEE International Conference on Rehabilitation Robotics *2011, 1-6.*

Schneider, S., Schönle, P.W., Altenmüller, E. & Münte, T. (2007). Using musical instruments to improve motor skill recovery following a stroke. Journal of Neurology, *254; 10, 1339-46.*

Thaut, M.H., Kenyon, G.P., Mc Intosh, G.C. & Hoemberg, V. (2002). Kinematic optimization of spatiotemporal patterns in paretic arm training with stroke patients. Neuropsychologica, *40, 1073–1081.*

Thibaut, A., Chatelle, C., Ziegler, E., Bruno, M.A., Laureys, S. & Grosseries O. (2013). Spasticity after stroke: Physiology, assessment and treatment. Brain Injury, *27; 10, 1093- 1105.*

Whitall, J., Waller, S.M., Silver, K.H.C. & Macko, R.F. (2009). Repetitive bilateral arm training with rhythmic auditory cueing improves motor function in chronic hemiparetic stroke. Stroke, *31, 2390–2395.*

TOWARDS A VOCABULARY FOR DESCRIBING 3D MOTION DATA IN FUNCTIONALLY ORIENTED MUSIC THERAPY

Loreen Pogrzeba[1], Margareta Ericsson[2], Karina Larsson[2], Markus Wacker[1], Bernhard Jung[3]
[1]University of Applied Sciences Dresden, Faculty of Information Technology and Mathematics, Germany, [2]Fmt Behandlingscenter Eskilstuna Ab, Sweden, [3]Tu Bergakademie Freiberg, Faculty of Mathematics and Computer Sciences, Germany
Time: 10:00

ABSTRACT

Functionally Oriented Music Therapy (FMT) is a special kind of active music therapy, which is focused on the neuropsychological development of patients, evaluated by the observation of the patients' motor skills and their ability to perceive and react to their environment. The rehabilitation progress is interpreted with the help of observation criteria like stability, hand function, and cross motion. In a field study we recorded three-dimensional motion data of patients with stroke and Parkinson's disease during their FMT therapy sessions. The results facilitated the development of a computer-assisted software tool supporting FMT therapists with the evaluation of their patients' rehabilitation progress. Accordingly, we transferred the FMT observation criteria, so far merely based on the personal experience of the therapists, to measurable and

calculable parameters to strengthen the assessment of the therapist. The following paper will give an insight into this transfer process which was completed during the first phase of our study. The measurability of the therapy progress permits a long-term computation of the effectiveness of the whole FMT therapy approach for the first time, and the extension of FMT observation criteria with motion analysis parameters leads to a holistic view on FMT. It enables the collaboration between therapists, patients, and physicians, due to a joined assessment vocabulary. The long-term computer-assisted observation is an important contribution to explore the effectiveness of FMT and an important step towards future personalized medical care in this working field.

1. INTRODUCTION

In 1975 the music teacher Lasse Hjelm started to develop his neuromuscular therapy approach called Functionally Oriented Music Therapy (FMT). In this non-verbal therapy the therapist uses music as tool to communicate with the individually treated person, the so-called adept. Music aims to catch the adept's attention and motivates him emotionally to perceive, react, move, play, and interact. So FMT is applicable for all persons regardless of age, and is often applied to children with learning difficulties and retarded development as well as persons with neurologic or motoric diseases, e.g. Parkinson's disease or stroke. FMT merges music, rhythmical motion and breathing, healthy seating, standing and walking, bodily exercises for improved flexibility, and self-confident communication in one structured working method. This variety renders this therapy approach and its assessment parameters very valuable also for other disciplines.

As FMT refers to a tradition of over thirty years and is based on the experience and knowledge of the practicing therapists, no official assessment standards exist. The assessment parameters are derived from the main publication of Hjelm (1995), who has established FMT as therapeutic approach. As far as we know only one English publication exists about assessment methods in FMT by Persson and Smideman (2002), which defines evaluation parameters in FMT and makes them accessible for the international community. Thus many inconsistent parameters are defined depending on the various training schools for FMT, and on the therapist's practical experience. This permissive definition may result in an individual interpretation of the traditional FMT assessment criteria in daily use which diversifies the specification of the criteria. Thus the range of the rank system classifying the scale of the functional problems of an adept is different and also the estimations which functional problems match which rank on the scale. As consequence objectivity in FMT assessment cannot be guaranteed. Furthermore these criteria are only one documental instrument to describe the progress of the treated person. Due to their experience FMT therapists recognize the progress of the adepts based on video recordings of the therapy sessions, hence feeling no need to carry out the evaluation in a consistent way. The collection of the assessment parameters is rather a supplement to the much more meaningful videos recordings. Indeed FMT therapists still need to struggle for acceptance of their therapeutic approach, and they are often encountered with skepticism by physicians and neurologists. With the advent of marker-free motion sensors they seize their chance to target a more objective, computer-assisted evaluation and define additional, objectively verifiable assessment parameters concordantly.

Since we started a binational field study about motion analysis in FMT in 2012 we have been facing the task to find a common language to get our knowledge and ideas across to each other, from therapist to computer scientist and vice versa. Thus we needed to find a kind of vocabulary to extend the vaguely defined and not standardized FMT observation criteria with very clear and measurable motion parameters for subsequent motion analysis algorithms. The following paper illustrates this process by offering a current definition of FMT therapeutic tools and explaining their transfer into motion analysis. It is complemented by a description of motion studies in music therapy and drumming, and closes with a discussion of future possibilities in research and analysis of motion patterns in FMT.

2. ASSESSMENT IN FMT

FMT is based on a holistic concept considering the concurrence of various bodily functions, such as perception, movement patterns, body control, breathing and concentration. Corresponding to its base assumption neurological disorders are reflected in a patient's behavior. The disorder could be a result of e.g. birth injuries, mental retardation, and neurologic diseases and often manifests itself in a visible anomaly or irregularity of the patient's motion. In FMT this visible deviation is called a functional constraint, because it reflects an impaired brain function: the greater the functional constraint, the lower the functional level of the patient. To observe, follow, rank, and finally treat the severity of functional problems the observation of motion is embedded in a system of specially designed observation tools, namely FMT codes and observation criteria.

2.1. FMT Codes

In a therapy session with a length of 15 to 30 minutes the therapist plays the piano while the adept plays the drums, cymbals, or various wind instruments. The arrangement of the percussion instruments is modified several times during the session, depending on the adept's functional constraints and his reactions on the supplied drum setup. Every arrangement of instruments is connected to a specific piano melody and a desired motion pattern. The motion pattern describes the sequence and order of strokes the adept has to perform, e.g. a symmetrical sequence from the center to the outer drums, or a stroke order from left to right. The combination of instrument arrangement, stroke order and melody is called a code. Within each code the therapist can vary the setup of the instruments slightly. These different versions of the same code increase the learning effect and allow for a more accurate and substantiated assessment of the functional level of the adept.

For all codes custom-made drumsticks and percussion instruments with adjustable individual stands are used. The codes can be performed in a sitting or standing position. Even special codes with longer walking distances between the drums are used. Different chairs with and without back- and armrests, exercise balls, balance pads, and cushions are also applied in order to challenge the adept's bodily skills. In addition to this technical equipment, the therapist aims to create situations that give the adept the opportunity to develop and provoke independent thinking as well as to enhance the ability to take initiative. The adept decides whether and how he follows the nonverbal signals of the therapist. His decisions are based on his capability and will to perceive and process nonverbal instructions and to react on them by means of motion, more precisely drumming or blowing.

Codes are varied during the session to practice several bodily functions and to target different cerebral regions. They are designed to address symmetrical or separate lateral motion and to improve the functions of several body parts and cognitive processing. The codes are strongly tied to the observation criteria in order to evaluate the adept's action. We will now explain these observation criteria in detail.

2.2. Observation Criteria

FMT therapists rely on different criteria to describe the bodily functions of their patients. These criteria tend to describe and evaluate neurologic functions of a patient by focusing on his bodily behavior, expressed by motion, and his ability to respond to external stimuli. So the observation criteria examine brain functions like perception, logic, memory, motor control, motor learning, and motor development. The assessment is based on the following 15 criteria, illustrated in Figure 1.

Model/logic. Model and logic pays attention to the ability of the adept to find a structure within an instrument arrangement. A structure is a repeating drumming pattern, e.g. a pattern in Western reading direction from left to right or a clockwise or anticlockwise rotational pattern. The patient should be able to derive the pattern without verbal instruction, just by trial-and-error or experience.

Perception. Perception includes auditory, visual, vestibular, tactile, and proprioceptive perception: the more impaired the overall perception, the worse the patient's ability to comprehend, distinguish, localize, disregard, and interpret the amount of incoming stimulation.

Breathing coordination. Breathing coordination is related to the ability to blow into a wind instrument and how the adept performs this action, e.g. weak or hard, in or out of rhythm with the piano. Since the therapist provides several wind instruments one after in successive order, he also analyzes the adept's capacity to reach the instruments, switch hands to receive and return them and his motoric planning of this repetitive action. In addition, breathing patterns such as rapid, slow or deep breathing or hyperventilation are registered, as well as the shape of the lips and mouth position during blowing.

Trunk rotation. Trunk rotation characterizes the ability to rotate the trunk within the horizontal plane, e.g. from left to right transverse rotation. According to Hjelm (1995) the human body consists of two functional systems: the lower system, consisting of feet, legs and pelvis, and the upper system, established by the torso, shoulders, neck, head, and arms. The two systems cooperate but also need to function separately, independently of each other.

Stability. Stability deals with the adept's ability to move in a balanced and safe way in the course of his sitting, standing, and walking action. This includes the skill to sense the floor or different underlays, like balance cushions or pads.

Hand function left and right (L/R). The development of the hands allows conclusions to the brain development, because a child learns and experiences the environment with the help of its hands. When the hand motion is limited, the brain is also not fully developed (Hjelm, 1995). Thus hand function L/R focus on the hand development in relation to the biological age, e.g. the way the drum sticks are grasped and gripped. Motor development in general grasping, reaching and postural control is described by Haywood, Roberton and Getchell (2012).

Wrist function L/R. Wrist function describes the mobility of the wrists, hence whether the motion is executed in a stiff and unstable or flexible and soft way. Stiff motion is manifested in restricted range of motion (ROM), unstable motion in increased ROM in certain directions.

Foot and hand coordination. Foot and hand coordination deals with coordinated, simultaneous and rhythmical motion of hands and feet.

Foot L/R. The therapist pays attention to the motion of the feet while the adept is playing a feet drum. He observes whether he is able to rhythmically stomp and how he is using his feet and ankles to establish a stable sitting position.

Total coordination. The above mentioned criteria are summarized as total coordination. It considers and summarizes the motion patterns and coordination of all body parts as well as the patient's perception, stability, logic, and breathing.

Separate lateral motion. Separate lateral motion reflects the adept's ability to move his left and right body half rhythmically in a different drumming pattern.

Side difference. Side difference examines the execution of motion of the left and right half of the body, especially when the left- or right-handed adept shows an unusual dominance of one hand**.** Dominance can be manifested so strongly, that the adept does not use the non-dominant hand during single-handed drumming or grasping. In addition the therapist focuses on a concordant speed of the left and right hand in both-handed drumming.

Cross motion. The therapist assesses the adept's skill to overcome the median line with his hands and arms. To execute this motion the brain hemispheres need to synchronize and adopt body control by turns.

Interaction. Interaction describes the ability of the adept to react non-verbally to the therapist, to handle the musical equipment and to orientate and move within the operating space spanned by the instruments.

The criteria mentioned from model/logic to foot L/R are related to the motoric and cognitive abilities of the adept. The observation criteria side difference, separate lateral motion, and cross motion refer to an imaginary center line called median line, dividing the adept's body in a left and right half. Finally, interaction deals with the communication and contact to the therapist.

We now shortly consider the above individually mentioned observation criteria in relation to each other using the example of trunk rotation: A wide perception is based on a good functional level of the trunk to turn the upper body towards points of interest. In cases of reduced trunk flexibility, stability is impaired, because a mobile trunk is necessary to tare the center of mass. Side difference is increased with less developed trunk rotation, because the impairment passes through the anatomic chain down to the hands and feet. Separate lateral motion and cross motion are also restricted by reduced mobility in the trunk. However, the variously notable relationships and dependencies between the individual observation criteria complicate the evaluation considerably.

Finally, the quality of the patient's response in all observation criteria is rated on an ordinal scale which scores his functional problems in a range from 0 - no functional problems to 6 - severe functional problems. The performance of the adept is assessed within a chosen time interval by collecting the ratings for each observation criterion and comparing them with each other. Up to now the therapists are using video recordings to document the progress of the adepts. To allow a computer-assisted evaluation in the future we need to transfer their observation criteria into measurable motion parameters.

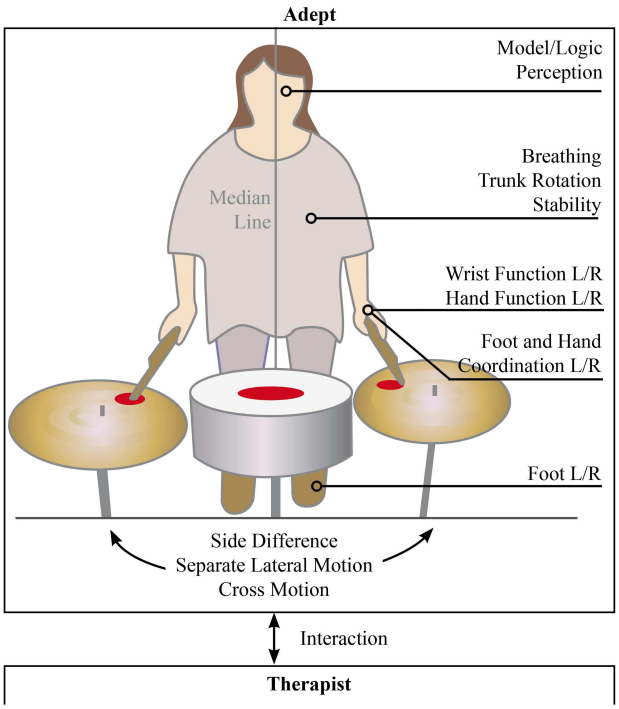


Figure 1: Schematic illustration of FMT observation criteria, except for total coordination

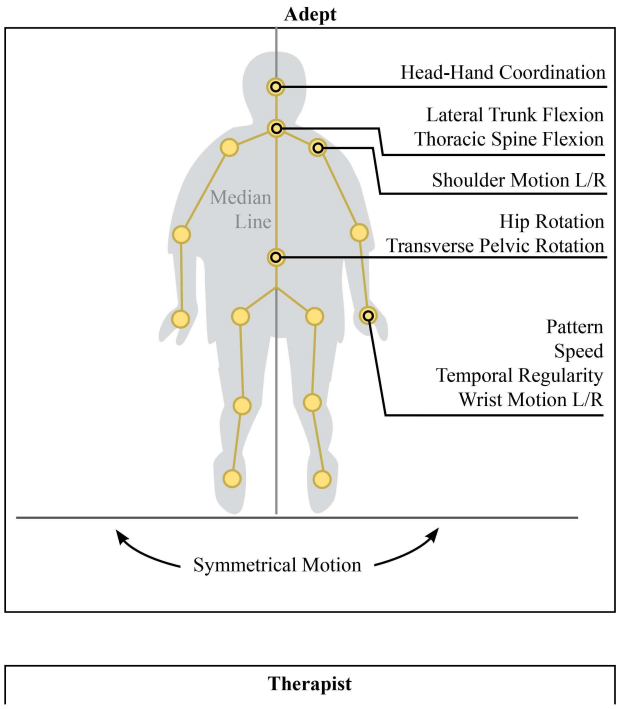


Figure 2: Schematic illustration of FMT 3D motion data vocabulary

3. Transfer of FMT Criteria to Motion Parameters

With the advent of low-cost hardware for motion recordings new possibilities in medicine and therapy emerged using motion data to evaluate the patient's progress. With the appearance of digital technical devices to measure and record motion, motion analysis became a possibility to examine, compare and classify motion. In our field study, we recorded motion of ten persons with Parkinson's disease and ten Stroke patients over a period of six months. A healthy control group was not involved. As recording device we used a Microsoft Kinect, which tracks the positional and rotational data of 15 joints of the adept during the FMT sessions. The resulting recordings permit for the first time to analyze the long-term therapeutic effects of FMT and facilitate new ways of presenting therapeutic progress for patients, physicians, and the therapists themselves. To render this motion data and progress depiction understandable and comprehensible for these very inhomogeneous target groups it was necessary to transfer the FMT observation criteria described in the previous section into quantitative and measurable motion parameters. In the following we would like to summarize motion studies and assessed motion parameters in music therapy and drumming. We will then present our developed 3d motion data vocabulary for FMT.

2.3. Motion Analysis in Music Therapy

Various studies in the fields of music therapy approached motion analysis. Hereinafter we would like to give an overview to motion studies aiming to analyze and evaluate the effects of music therapy or cyclic motion patterns, e.g. drumming.

Burger, Ala-Ruona and Järvinen- Lepistö (2012) traced improvements of the vertical positions of wrist and finger at an easy and difficult drumming task in FMT, illustrated in graphs and motion trajectories. Since this was a preliminary approach, a final analysis of the motion study with 45 stroke patients is still pending. In the field of neurological music therapy (NMT) Thaut (2005) gave insights in the use of Rhythmic Auditory Stimulation (RAS) to successfully treat patients with neurologic disabilities in gait. NMT also uses Therapeutic Instrumental Music Performance (TIMP), providing parallels to FMT. According to Chong, Cho, Jeong, and Kim (2013) TIMP also uses percussion instruments for the training of gross motor skills and – like FMT – it relies on different playing methods and changes in playing posture to invoke and strengthen motor patterns, "e.g., placing various sizes of drums at different heights for the player in order to induce expanded range of arm movement" (p. 421). They executed a study with 5 adults suffering from cerebral palsy (CP) assessing fine motor skills during piano play. The participants performed a repetitive play of preferred musical notes at two sessions a week for nine weeks at the maximum. In comparison to a healthy control group the CP patients slightly improved the velocity of key pressing force of the affected hand, most significantly for the index and small fingers. Schneider, Schönle, Altenmüller, and Münte (2007) found improvements in the range, speed and quality of movements of 20 stroke patients after 15 sessions of individual musical training with drum pads and a MIDI-piano. They measured the number of full cycles per second, the number of inversions of velocity and the average maximum angular velocity during whole hand tapping and index finger tapping for each hand before and after the whole treatment. They combined motion analysis with motion tests, e. g. Action Research Arm Test, and assessed improvements in everyday activities. In a follow-up study they extended their approach with EEG recordings (Altenmüller, Marco-Pallares, Münte, and Schneider, 2009).

In a motion study with four healthy drummers, Dahl (2004) investigated connections between striking velocity, preparation height and timing for different tempos, dynamic levels and striking surfaces. With the help of a 4-marker-setup and a cylindrical force plate she recorded the percussionists performing right-handed, cyclic drumming patterns with accents on every forth note. She found a close correlation between

striking velocity and preparatory height of the drumstick, especially greater heights at accented strokes, and rising striking velocity with increasing dynamic level and on soft and normal surfaces. Concerning timing she traced a small drift in tempo for all players. Similar studies were conducted by Dahl, Grossbach and Altenmüller (2011a, 2011b) in slightly varied realization, e.g. extended with audio recordings, the measurement of the contact duration and with changed marker setup and tempi. Moreover Dahl, Grossbach and Altenmüller (2011b) considered motion recordings of two right-handed drummers with focal dystonia of the left arm. They discovered increased time variability at the fastest tempo of 300 beats per minute in comparison with the performance of the healthy unaffected arm and lower peak acceleration in both-handed drumming compared to healthy percussionists. Time variability and motion coordination at slow and intermediate tempi were not affected. In addition to motion analysis, rating scales and assessment tests provide methods to evaluate the bodily skills of patients. They potentially serve as tools to state the need for treatment and to estimate the effectiveness of therapy based on clearly defined exercises or items. Due to their dependence on the patient groups and application field we refer to the "Handbook of Neurologic Rating Scales" by Herndon (2006) and the Rehabilitation Measures Database (http://www.rehabmeasures.org) for assessment of patient groups similar to these in FMT.

Although approaches for motion analysis differ depending on the type of therapy and the used parameters to investigate given research hypotheses, we presented motion studies which also deal with motion analysis in music therapy or cyclic motion patterns conducted with sound. In the following chapter we introduce our 3d motion data vocabulary for assessment in FMT.

2.4. 3D Motion Data Vocabulary in FMT

In FMT, motion analysis opens up new methods to arrange, classify, compare and visualize the bodily performance of the adepts. In a team of FMT therapists and computer scientists we reassembled the traditional observation criteria explained above to new meaningful and at the same time measurable parameters. In the following we will present our assessment vocabulary in FMT as well as the preconditions for the transfer of FMT criteria to 3d motion parameters and provide impulses for further research.

To transfer the observation criteria into motion data vocabulary we encounter three basic preconditions concerning the preparation, realization and evaluation of the motion recordings. In the preparation phase, limitations occur due to the chosen recording technology. Here we focused on criteria observable via motion data, whose parameters – the number of recorded joints, the accuracy and the frame rate – are determined by the used device. Additionally, the device was required to be light-weighted, mobile, low-maintenance, of robust construction and low-priced. In the realization phase more conditions arose from the concept of the FMT approach. All motion parameters needed to fit unobtrusively into the therapy sessions. They had to be recordable without voice instructions and needed to smoothly integrate into the FMT codes with percussion or wind instruments without requiring additional actions or behavior that would distract either patients or therapists. Finally in the evaluation phase, the motion data parameters demanded for intuitive and quick comprehension by therapists in their daily assessments. The analogies to their known observation criteria had to be clear, extensions and omissions obvious.

Under these preconditions motion analysis allows only a partial evaluation of all FMT observation criteria. We have chosen 6 out of 15 FMT criteria for a translation into motion analysis parameters: model/logic, trunk rotation, stability, wrist function L/R, side difference, separate lateral motion and cross motion. In the following we present our 3d motion data vocabulary in FMT, illustrated in Fig. 2, as well as the connection to the traditional FMT observation criteria.

Pattern. Pattern covers the FMT observation criteria model/logic, separate lateral motion and cross motion. It implies the appearance of a repetitive motion pattern performed by the adept. Detected repetitive patterns are compared to a FMT code catalogue containing the different associated beat sequences. If the adept performs a FMT code correctly in accordance to the intended stroke sequence, the motion pattern is recognized and associated with that FMT code. Via this detection the pattern provides information on the adept's ability to perform separate lateral motion and cross motion, which is necessary to execute several FMT codes.

Symmetrical motion. Symmetrical motion refers to the observation criterion side difference. It reflects the accordance of the motion of the left and right body half, especially of the hands and wrists. Spatial symmetry is expressed by a coincident motion trajectory of the involved joints of the left and right body half, and congruent joint angles. During several repetitions, the motion ideally experiences little spatial variance and outliers. Especially the contact points of the drum sticks with the percussion instruments should be located in a narrow spatial areal in the center of the instrument.

Temporal regularity and speed. Temporal regularity and speed is related to the time-dependent aspects of side difference. Since the adept himself determines the speed at which he performs music codes, his chosen tempo is of high importance. Consequently temporal regularity and speed represent time-dependent motion parameters. Temporal regularity expresses the adept's ability to keep his chosen pace during the performance of one code and its versions. Pauses within a code, time-lags between the hits of the left and right drum stick, additional or missing beats and abortion of the code disturb temporal regularity. Speed describes the average drumming tempo of an adept in cycles per second (depending on the chosen FMT code and its involved number of instruments) and beats per minute. A faster or slower execution in comparison to a preferred or average drumming speed is a sign of fatigue or particular effort due to a higher degree of difficulty of the demanded drumming pattern. Parkinson patients often prefer a slower, children with ADHD a faster speed in the first FMT therapy sessions in comparison to the end of their treatment.

Ergonomic posture. Ergonomic posture specifies the FMT observation point stability and consists of four parts. It covers lateral flexion at the trunk, thoracic spine flexion, hip rotation around the longitudinal axis and shoulder motion during sitting. The therapist strives for an upright sitting position with no lateral flexion of the trunk. This neutral spine posture is ideally reflected in a straight upper back without any outward rounding of neither the thoracic spine area nor a postural kyphosis. The legs should be in a neutral uncrossed position with full contact of the feet to the floor and without any internal or demonstrative external hip rotation. Postures with the arm positions fixed close to the trunk need further treatment, because the adept is not able to keep a stable sitting position during flexible arm motion. The adept should be able to flex, extend, adduct, and abduct his shoulders to reach all the instruments within his reaching area. Ergonomic considerations are currently not being taken into account for standing and walking action, since some patients are using wheelchairs or walking frames and the majority of FMT codes are performed in a sitting position, depending on the functional problems of the adept.

Transverse pelvic rotation. Transverse pelvic rotation is directly transferred from the FMT observation criterion trunk rotation. It describes the left and right transverse pelvic rotation. An apparently reduced ROM or a motion bias in one direction demands further treatment.

Wrist motion. Since all instruments must be grasped, guided, and played with the hands, the assessment of the wrist function L/R is highly important. Wrist motion focuses on the ROM of the hands, especially flexion, extension and pro- and supination. During the drumming performance none of the motion angles should remain unchanged as this would be an indication for stiffness. Additionally, reduced flexion and hyperextension of the wrists adversely affect the motion, because in this case stronger arm and trunk motion is required in order to reach the drums with the drum sticks.

Head-hand coordination. This parameter is related to the mobility of the neck and the cervical spine and tries to cover partly the FMT observation criterion perception. It can be viewed as a simplified motion parameter for eye-hand coordination. An unrestricted motion at the neck is desired, where the head and the eyes freely follow the motion of the hands. A stiff head position without any motion at the neck, and a motion with remaining lateral bend or neck extension needs further treatment.

The remaining FMT criteria have not been involved into the 3d motion data vocabulary and open up opportunities for future research. Perception is difficult to assess quantitatively, because it is a compound factor of several sensations requiring many devices for the recording of all incoming stimuli. For continuous computer-assisted evaluation in FMT it would require too much effort. While we have omitted to collect data on the adept's breathing coordination, there are now methods available to unobtrusively document the breathing rhythm with the use of a laser-based active triangulation sensor (Bauer et al., 2012) or a Microsoft Kinect sensor (Noonan et al., 2012). In addition, finger tracking SDKs are available, allowing for computer-assisted assessment of the finger position and grasping gestures in the future and permitting analysis of hand function L/R. However Robertson, Vink, Regenbrecht, Lutteroth and Wünsche (2013) currently criticize the low tracking performance of finger tracking SDKs on a mid-range computer without a graphics card. We here point out that portable computers with on-board graphics are the devices predominantly used by FMT therapists. The robust recording of rhythmic feet motion due to the observation criterion feet L/R as well as foot and hand coordination highly depends on the position of the Microsoft Kinect sensor. During the recording sessions, we focused on a stable tracking of the upper body and an unobtrusive, static attachment of the sensor to the ceiling in the therapy room. This setup led to missing motion data of the feet and knees, because they were mainly covered by drums and could not be tracked continuously. In future recordings, the detection of rhythmical motion of the feet might be derivable from a pressure sensor attached to the foot pedal of a drum or an in-shoe pressure measurement device. Furthermore, affordable pressure sensors could extend motion analysis concerning sitting ergonomics and body weight distribution. In addition, some information about interaction are probably be reproducible from the Microsoft Kinect data, using recognition of facial expressions and analysis of the person traces following DeCamp, Shaw, Kubat and Roy (2010). Easy recognition of facial expressions, heart rate and better joint recognition, e.g. tips of hands and thumbs, are amongst others announced for the Microsoft Kinect version 2.0 (Heddle, 2013; Wired, 2013).

6. CONCLUSIONS AND FUTURE WORK

In the previous chapters we presented the FMT therapy approach, its working methods, and assessment criteria. With the transfer of the FMT criteria to known and understandable motion parameters we have established a scientific basis to objectively assess and evaluate the therapy progress and created an additional unifying assessment vocabulary for FMT therapists, physicians, and patients. We introduced this type of neuromuscular therapy to the general public, especially to physicians and neurologists. Its structured design builds on codes and its multifunctional assessment features, involving sitting ergonomics, rhythm, logical patterns, and mobility of the pelvis and wrists, make it highly valuable for patients with impaired brain and bodily functions. Additionally, healthy people are invited to challenge themselves and practice their postural control, separate lateral motion, and ability for logical deduction in a playful, motivating manner in FMT sessions.

Currently the development of computer-assisted motion analysis in FMT is limited to the motion parameters introduced in this paper, and their possibilities for comprehensible and clear visualization. In the beginning of our research, we agreed to deliberately restrict the number of motion parameters for reasons of clarity and comprehensibility in documentation and visualization. Further advantages consist in the facilitated familiarization with the future software and the lower expenditures in technical equipment, required disc space and working memory. However, it is still questionable whether the accuracy of the Microsoft Kinect is accepted by all target groups for all defined motion parameters. Otherwise, marker-based methods need to be provided supportively to prove the efficiency of FMT. In addition, a combination of our motion parameters with an established assessment method could be an advantage for a well-grounded study concerning the efficiency of FMT. It could allow conclusions about to what extent the measured physical changes affect the adepts' daily life and everyday actions like eating or dressing in a positive way. Thus this assessment vocabulary is a first approach for scientific discourse, which has to prove itself in practice.

We will continue with the analysis of the motion data following the above mentioned assessment vocabulary to answer the question of the efficiency of the FMT therapy approach. Our analysis of data samples indicates improvements of the adepts' functional levels, so we are eager and confident to refine our assessment approach for future application in daily practice.

7. ACKNOWLEDGMENT

This project was partially funded by the European Commission (European Social Fund). Advice given by Dr. Anita Granberg, Ulla Sterner and Mikael Walldin has been a great help in transferring the FMT observation criteria into motion parameters. Thanks to all participants who took part in this project. We cordially thank all the FMT therapists who have inspired us with their ideas and feedback.

8. REFERENCES

Altenmüller, E., Marco-Pallares, J., Münte, T. F., & Schneider, S. (2009). Neural reorganization underlies improvement in stroke-induced motor dysfunction by music-supported therapy. The Neurosciences and Music III - Disorders and Plasticity: Ann. N.Y. Acad. Sci., 1169, *395–405. doi: 10.1111/j.1749-6632.2009.04580.x*

Bauer, S., Berkels, B., Ettl, S., Arold, O., Hornegger, J., & Rumpf, M. (2012). Marker-less reconstruction of dense 4-D surface motion fields using active laser triangulation for respiratory motion management. Med. Image. Comput. Comput. Assist. Interv., 15 (Pt 1), *414–21. doi: 10.1007/978-3-642-33415-3_51*

Burger, B., Ala-Ruona, E., & Järvinen-Lepistö, P. (2012). Employing motion capture in studying the effects of active music therapy for post-stroke recovery. In 7th Nordic Music Therapy Congress (NMTC). *Jyväskylä, Finland. Poster.*

Chong, H. J., Cho S.-R., Jeong, E., & Kim, S. J. (2013). Finger exercise with keyboard playing in adults with cerebral palsy: A preliminary study. J. Exerc. Rehabil., 9(4), *420–25. doi: 10.12965/jer.130050*

Clark, R. A., Pua, Y.-H., Fortin, K., Ritchie, C., Webster, K. E., Denehy, L., & Bryant, A. L. (2012). Validity of the Microsoft Kinect for assessment of postural control. Gait & Posture 36, *372–77. PubMed PMID: 22633015.*

Dahl, S. (2004). Playing the accent - comparing striking velocity and timing in an ostinato rhythm performed by four drummers. Acta Acustica united with Acustica, 90(4), *762–76. Retrieved from http://www.speech.kth.se/*

Dahl, S., Grossbach, M., & Altenmüller, E. (2011a). Effect of dynamic level in drumming: Measurement of striking velocity, force, and sound level. Proc. of Forum Acusticum. Danish Acoustical Society, *621–24. Retrieved from http://vbn.aau.dk*

Dahl, S., Grossbach, M., & Altenmüller, E. (2011b). Good playing practice when drumming: Influence of tempo on timing and preparatory movements for healthy and dystonic players. Proc. of Intern. Symp. on Perfor. Sci. 2, *237–242. Retrieved frm http://vbn.aau.dk/*

DeCamp, P., Shaw, G., Kubat, R. & Roy, D. (2010). An immersive system for browsing and visualizing surveillance video. ACM Multimedia, *371–80.*

Fernández-Baena, A., Susín, A., Lligadas, X. (2012). Biomechanical validation of upper-body and lower-body joint movements of Kinect motion capture data. In Proc. 4-th Int. Conf. on Intelligent Networking and Collaborate Systems, *656–61 doi: 10.1109/iNCoS.2012.66*

Haywood, K. M., Roberton, M. A., & Getchell, N. (2012). Advanced analysis of motor development. *Human Kinetics.Heddle, B. (2013, May 23). The new generation Kinect for windows sensor is coming next year. Retrieved from http://blogs.msdn.com/b/kinectforwindows/archive/2013/05/23/the-new-generation-kinect-for-windows-sensor-is-coming-next-year.aspx*

Herndon, R. M. (Ed.). (2006). Handbook of neurologic rating scales *(2nd ed.). New York: Demos Medical Publishing.*

Hjelm, L. (1995). Med musik som medel: FMT-metoden, som den blev till … *Uppsala: Musikterapiinstitutet Uppsala.*

Noonan, P. J., Howard, J., Tout, D., Armstrong, I., Williams, H. A., Cootes, T. F., Hallett, W. A., & Hinz, R. (2012). Accurate markerless respiratory tracking for gated whole body pet using the Microsoft Kinect. IEEE Nuclear Science Symp. and Med. Imaging Conf., *3973–74. doi: 10.1109/NSSMIC.2012.6551910*

Persson, L., & Smideman, G. (2002). Terms and criteria for assessment and evaluation with a neuropsychological perspective. Conf. Proc. of the 10th World Congress of Music Therapy, *1283–98.*

Robertson, C., Vink, L., Regenbrecht, H., Lutteroth, C., & Wünsche, B. C. (2013). Mixed reality Kinect mirror box for stroke rehabilitation. Conf. Proc. Image and Vision Comput New Zealand, 231–35. doi: 10.1109/IVCNZ.2013. 6727021

Schneider, S., Schönle, P. W., Altenmüller, E., & Münte, T. F. (2007). Using musical instruments to improve motor skill recovery following a stroke. J Neurol., 254(10), *1339–46. doi: 10.1007/s00415-006-0523-2*

Thaut, M., H. (2005). Physiological and motor responses to music stimuli. In R. F. Unkefer, & M. H. Thaut (Eds.), Music Therapy in the Treatment of Adults with Mental Disorders (pp. 33–41). Gilsum: Barcelona Publishers.

Wired (2013, May 20). New Xbox One- Kinect: Exclusive Wired Video [Video file]. Retrieved from http://video.wired.com/watch/game-life-new-xbox-one-
kinect-exclusive-wired-video

[9C] Musical Ability

MONTREAL 09:00- 10:30 Friday 8 Aug 2014

MUSICAL ABILITY TO SEARCH FOR ESSENTIAL MEANING IN AN APPRECIATION TASK -THE CASE OF MINIMAL MUSIC-

Yuko Odahara[1], Yumi Hoshino[2]
[1]Hyogo University of Teacher Education, Japan, [2]Okayama University, Japan

Time: 09:00

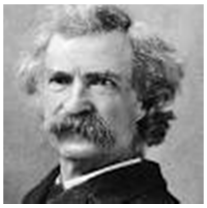
The musical guidelines in Japan's National curriculum strongly recommend connecting language and music, and having children verbalize and explain the music they hear and how they feel. In order to be able to verbalize a musical piece, classes which approach the essence of a musical piece to help children appreciate and understand, must be given. However, can children grasp detailed composition techniques, and can they approach the essence of a musical piece? How and what should they be made to observe in a musical piece, in order to approach its essence? To raise awareness of these issues, we conducted an experiment that used real and imitation minimal music for junior high school students. The subjects were 155 seventh graders. They took a pre-test, classes, and post-test. Students had three classes about appreciation and composition. Four pieces of real and imitation minimal music were prepared for classes. Additionally, twelve pieces of real and imitation minimal music were prepared for pre-test and post-test as stimuli. The imitation music consisted of our own musical pieces. They were played as stimuli at random for thirty seconds each time. Students were required to distinguish real or imitation while listening to each stimulus. In the pre-test ($p < .001$), musical pieces which the majority of students distinguished correctly were seven of twelve pieces. In the post-test ($p < .001$), they were eight of twelve pieces. The results of the experiment should that: (1) Students based their distinction on the superficial parts of the musical piece in the pre-test. (2) Students distinguished the musical pieces by the structure of the music in the post-test. (3) Students could find differences between real and imitation minimal music in the post-test. These results suggest that it is effective to use real and imitation minimal music to approach the essence of musical pieces.

REDEFINING MUSIC EXCELLENCE: MUSICAL FUTURES IN SPECIAL AND INCLUSIVE EDUCATION.

Helen Farrell[1]
[1]The University of Melbourne, Australia

Time: 09:30

ABSTRACT



"... twenty years from now you will be more disappointed by the things that you didn't do than by the ones you did do ... so throw off the bowlines ... sail away from the safe harbor ... catch the trade winds in your sails ... explore ... dream ... discover ..." Author, Samuel Langhorne Clemens (1835-1910), better known by his pen name Mark Twain.

I am quite sure that most of us have enjoyed Twain's *The Adventures of Tom Sawyer* (1876) and its sequel, *Adventures of Huckleberry Finn* (1885). But, what can we learn from this, one of his many quotations, in what ought to be a continuous journey to learn more and more about assessment and reporting of achievement of diverse musical development and learning.

Categories of trends and issues impacting on special and inclusive education are numerous. For example, nobody can truly know what is locked in the world of individuals with the many classifications of complex special needs.

The paper reviews some of the neuroscientific research that informs understandings of fundamental processes of musical development and learning.

Music neuroscience is a rapidly maturing field. Recent research highlights the plasticity of neural systems that underpin engagement with musical activity. Music neuroscience informs understandings of fundamental processes in musical development and learning; and in turn, offers sophisticated accounts of communication, physical, cognitive, and social/emotional behaviours. The author notes a profound paucity of such research with and for children and young people with complex special needs.

Many challenging, sometimes controversial, implications for sustainable politically, economically, culturally and socially public policy, and professional practice are encountered head-on in this circumstance, for example, teachers, parents, paramedical and medical professionals, academics, students, administrators, policy makers and others whose work impinge on curriculum development and implementation.

The author offers an improved understanding of the extraordinary complexities that encompass fair, valid, quality and credible rubrics in public standards-based music curriculum, assessment and reporting frameworks to contribute to a respectful redefinition of musical excellence in special and inclusive education.

An improved understanding of these extraordinary complexities will much reduce fear, unease and distrust. The phenomenon would seem logical.

1. THE MOST EXTRAORDINARY THINKING MUSICIAN AND MUSICAL THINKER THEY CAN BE.

Work with individuals with complex special needs is profoundly complex and challenging. The accumulation of the many decades of extraordinary evidence based research and scholarship can be explored to optimise innovative solutions (Australian Research Council, 1997; Creswell & Plano Clark 2010; Denzin & Lincoln (editors), 2011; Eisner, 1986; Eisner, 1998; Fielding & Lee, 1993; Fitzmaurice et al (editors), 2008; Kemmis & McTaggart, 1988; Miles & Huberman, 1994; Miles & Weitzman, 1994; National Health and Medical Research Council (NHMRC) 2007; Patton, 2002; Spradley, 1980; Stake, 1978; Stake, 1983; Stake, 1988; Stake, 1995; Thomas & Pring (editors), 2004); Yin, 2009; Yin, 2012).

It *is* possible to make such an apparently insular world fulfilling. Music seems to provide an outlet for expression and communication with others in an otherwise strange and confusing universe. Indeed, some may possess extraordinary musical gifts and talents despite sometimes severe, profound and multiple limitations in other domains (Doidge, (2010; Levitin and Bellugi, 1998; Miller, 1989; Ockelford, 2007; Ockelford, 2008; Ockelford in G. McPherson & G. Welch (editors) (volume 2), 2012a; Ockelford in G. McPherson & G. Welch (editors) (volume 2), 2012b; Sacks, 2007).

2. THE NEUROSCIENTIFIC RESEARCH.

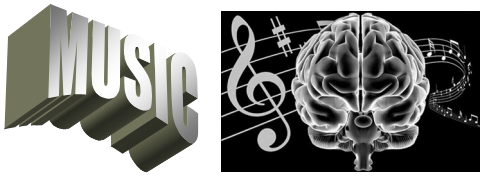
Recent neuroscientific research has suggested some of the most extraordinary interdisciplinary evidence based research and scholarship into the neurobiology, neurology and neuropsychology (Alberts, Johnson, Lewis, Raff, Roberts & Walter, 2008; Anderson, Notrham, Hendy & Wrennal, 2001; Andrewes, 2001; Bear, Connors, Paradiso, 2007; Blumenfeld, 2010; Gazzaniga, 2009; Kandel, Schwartz & Jessell (editors), 2013; Nelson & Cox, 2013; Schapira (editor), 2007).

Neuroscience has potential to significantly enhance understandings of development and learning processes; and implications for effective policy and professional practice in creating musical futures in *special and inclusive education*.



The interdisciplinary nature of neuroscientific research is rapidly developing. The nexus between these fields for education, paramedical, medical and other professionals is still in its infancy. Fruitful areas for collaboration are possible.

One of the only activities that activates,
stimulates and uses the entire brain is …



Music takes place in many contexts, both formal and informal. An extensive body of evidence based inquiry and research related to biological, neurological and psychological aspects of musical development and learning has accumulated over at least 40 years. Inquiry and research embrace notions of (1) musical performance, improvisation and composition; (2) the unique properties of musical elements (rhythm, melody, harmony, expression, timbre, style, form); and (3) affective behaviours and musical preferences within and across the life span from conception and infancy, through childhood and adolescence and into adulthood (Davies, 1978; Deutsch, (editor), 2012; Gruhn & Rauscher (editors), 2007; Hallam, Cross & Thaut (editors), 2009; Hargreaves, 1986; Hargreaves, Miell & MacDonald (editors), 2012; Hodges, (editor), 1996; Peretz & Zatorre, 2005; Radocy & Boyle, 2003; Thompson, 2009)

Auditory Cognitive Neuroscience (ACN) and cognate disciplines have demonstrated the very clear influence of music in its capacity to integrate multiple brain systems. For example, neuroimaging studies demonstrate that multiple brain networks are activated whilst *making music* and *responding to music* (Merrett and Wilson in N. S. Rickard & K. McFerran (editors), 2011).

3. CHALLENGES AND OPPORUNITIES: MUSICAL FUTURES IN SPECIAL AND INCLUSIVE EDUCATION.

Many challenges and opportunities remain for those who seek to move from the brain scan into effective recalibration of public policy and professional practice for teachers, parents, paramedical and medical professionals, academics, students, administrators, policy makers and others in special and inclusive education sectors (Althaus, 2007; Bessant, Watts, Dalton & Smyth, 2006; Fischer et al., 2007; Hill & Hupe, 2009; Maddison, 2009; McClelland & Smyth (editors), 2006; Mendes, 2008; Wanna, Butcher & Freyens, 2010).

Curriculum, assessment and reporting frameworks simply *must* take courage to accomplish innovative and sustainable outcomes that, in some modest ways, mitigate political, economic, social and cultural exclusion of children and young people of compulsory school age with the many classifications of complex special needs (Browder, et al, 2004; Department of Education, Training and Workplace Relations, 2009; Griffin, et al, 2010).

So, let's throw off the bowlines. Let's sail away from the safe harbor. Let's catch the trade winds in our [corporate education] sails. Let's explore. Let's dream. Let's discover.

Musical development and learning in children and young people with complex special needs *can* be promoted in fair, valid and credible public standards-based music curriculum, assessment and reporting frameworks.

Additional levels of knowledge, understanding and skills descriptors; achievement standards expected; and general cross-curriculum capabilities *can* be developed and implemented for public standards-based music curriculum, assessment and reporting frameworks.

Additional rubrics *can* be supported with descriptors of pre-intentional; intentional; early symbolic; and symbolic levels of development and learning.

Fair, valid and credible assessment of musical development and learning in these children and young people *can* be reported with additional levels of rubrics (Hattie & Timperley, 2007; Kellaghan, & Greaney, 2001; Masters, 2013).

Successful professional practices *can* be fairly, validly and credibly correlated to learning standards and pathways that musical development and learning in these children and young people is promoted with development of assistive and augmentative resources (Curriculum Corporation, 1994; Qualifications and Curriculum Authority (Department for Education and Employment), 2001).

Sustainable foundations for musical futures in special and inclusive education *can* be championed (Henley, 2011; James, S (editor), 2012; Office of Standards in Education (OFSTED), 2012a; Office of Standards in Education (OFSTED), 2012b; Pascoe e al, 2005).

Twenty years from now, I will be *very* disappointed if I cannot reflect on, and be much encouraged by very positive efforts to tackle these significant and inherent challenges and opportunities, and shape strategy to *enrich* diversity in musical development and learning, rather than be *constrained* by the things that are not done. I wonder how I will approach the same key questions at the 2034 ICMPC!! I look forward to presenting findings!!

4. ACKNOWLEDGEMENTS.

The author would like to thank support of the *Melbourne Neuroscience Institute* (MNI), the *Music, Mind and Wellbeing* (MMW) and the *National Music Therapy Research Unit* (NMTRU) teams in the University of Melbourne, Australia.

5. FUNDING.

No grant from any public, corporate or not-for-profit sector funding agency was received.

6. REFERENCES.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts. K., & Walter, P. (2008). Molecular Biology of the Cell *(5th edition). New York: Garland Science.*

Althaus, C (2007). Australian Policy Handbook *(4th edition). Sydney: Allen & Unwin.*

Anderson, V., Notrham, E., Hendy, J., & Wrennal, J. (2001). Developmental Neuropsychology*:* A Clinical Approach*. London: Psychology Press.*
Andrewes, D.G. (*2001).* **Neuropsychology*:* From Theory to Practice.** *Hove, United Kingdom: Psychology Press.*
Australian Research Council (1997). Australian Code for the Responsible Conduct of Research*. Retrieved from <http://www.arc.gov.au>.*
Bear, M.F., Connors, B.W., & Paradiso, M.A. (2007). Neuroscience*:* Exploring the Brain *(3rd edition). Philadelphia: Lippincott Williams & Wilkins.*
Bessant, J., Watts, R., Dalton, T. & Smyth P. (2006). Talking Policy*:* How Social Policy is Made*. Sydney: Allen & Unwin.*
Blumenfeld, H. (2010). Neuroanatomy Through Clinical Cases *(2nd edition). Sunderland, MASS: Sinauer Associates.*
Browder, D. M., Flowers, C., Ahlgrim-Delzell, L., Karvonen, M., Spooner, F. & Algozzine, R. (2004). The alignment of alternate assessment content with academic and functional curricula. The Journal of Special Education, *37(4), 211-223.*
Creswell, J. & Plano Clark, V. (2010). Designing and conducting mixed methods research *(2nd edition). Thousand Oaks, CA: Sage Publications.*
Curriculum Corporation (1994). Towards Band 1 in The Arts*:* A Curriculum Profile for Australian Schools*. Carlton, Melbourne: Curriculum Corporation.*
Davies, J. B. (1978). The Psychology of Music*, Hutchinson & Co., London.*
Denzin, N. K. & Lincoln, Y. S. (editors) (2011). The Sage Handbook of Qualitative Research *(4th edition). Thousand Oaks CA: Sage Publications.*
Department of Education, Training and Workplace Relations (2009). Early Years Learning Framework*:* Belonging, Being and Becoming*. Barton, Australian Capital Territory: Commonwealth of Australia. Retrieved from <http://www.deewr.gov.au/eylf>.*
Department of Prime Minister and Cabinet (2008). Australia 2020 Summit*:* Final Report*. Barton, Australian Capital Territory: Commonwealth of Australia. Retrieved from <http://www.australia2020.gov.au/docs/final_report/2020_summit_report>.*
Deutsch, D. (editor) (2012). The Psychology of Music *(3rd edition). Hillsdale, New Jersey: Laurence Erlbaum.*
Doidge, N. (2010). The Brain that Changes Itself*:* Stories of Personal Triumph from the Frontiers of Brain Science *(revised edition). Carlton North, Victoria, Australia: Scribe Publications.*
Eisner, E. W. (1986). "A Secretary in the Classroom". Teaching and Teacher Education, 2*(4), pp. 325-328.*
Eisner, E. W. (1998). The Enlightened Eye*:* Qualitative Inquiry and the Enhancement of Educational Practice*. Upper Saddle River NJ: Merrill.*
Fielding, N. G. & Lee, R. M. (1993). Using Computers in Qualitative Research *(revised and updated). London: Sage.*
Fitzmaurice, F. et al (editors) (2008). Longitudinal Data Analysis*. Boca Raton, FL: CRC.*
Gazzaniga, M. (2009). The Cognitive Neurosciences *(4th edition). Cambridge, MASS: MIT Press.*
Griffin, P., Woods, K., Coles-Janess, B. & Roberts, E. (2010). "Mining the gold: Assessing the students by ability, not disability". Teacher, *34-37.*
Gruhn, W. & Rauscher, F.H. (editors) (2007). Neurosciences in Music Pedagogy*. New York: Nova Science Publishers.*
Hallam, S., Cross, I., & Thaut, M (editors) (2009). Oxford Handbook of Music Psychology. Oxford: Oxford University Press.
Hargreaves, D. J. (1986). The Developmental Psychology of Music*. Cambridge UK: Cambridge University Press.*
Hargreaves, D., Miell, D & MacDonald, R (editors) (2012) Musical Imaginations*:* Multidisciplinary Perspectives on Creativity, Performance and Perception*. Oxford: Oxford University Press.*
Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77*(1), 81–112.*
Henley, D (2011). The Importance of Music*:* A National Plan for Music Education*. London: Department for Education/Department for Culture, Media and Sport. Retrieved from <http://publications.education.gov.uk>.*
Hill. M. & Hupe, P. (2009). Implementing Public Policy: *An Introduction to the Study of Operational Governance (2nd edition). London: Sage Publications.*
Hodges, D. A. (editor) (1996). Handbook of Music Psychology *(2nd edition). San Antonio TX: IMR Press.*
James, S (editor) (2012). An Extraordinary School*:* Re-Modelling Special Education*. Melbourne: ACER Press.*
Kandel, E.R., Schwartz, J.H., & Jessell, T.M. (editors) (2013). Principles of Neural Science *(5th edition). New York: McGraw-Hill.*
Kellaghan, T., & Greaney, V. (2001). Using assessment to improve the quality of education*. Paris: UNESCO.*
Kemmis, S. & McTaggart, R. (1988). The Action Research Planner*. Geelong, Australia: Deakin University Press.*
Levitin, D.J. & Bellugi, U. (1998). Musical Abilities in Individuals with William Syndrome. Music Perception, 15*(4), 357-389.*
Levitin, D.J. (2008). The World in Six Songs: How the Musical Brain Created Human Nature*. London: Dutton Publishers.*
Maddison, S. (2009). Introduction to Australian Public Policy Theory and Practice*. Cambridge, UK: Cambridge University Press.*
Masters, G. N. (2013). "Reforming Educational Assessment: Imperatives, Principles and Challenges". Australian Education Review, *no. 57. Camberwell, Victoria, Australia: Australian Council for Educational Research. ISBN: 9781742862026.*
McClelland, A. & Smyth, P. (editors). (2006). Social Policy in Australia*:* Understanding for Action*. Melbourne: Oxford University Press.*
Mendes, P. (2008). Australia's Welfare Wars Revisited*:* The Players, The Politics and the Ideologies*. Sydney: UNSW Press.*
Merrett, D and Wilson, S. J. (2012). In Rickard, N. S. & McFerran, K. (editors). Lifelong engagement in music*:* Benefits for mental health and well-being*. New York: Nova Publishers.*
Miles, M. B. & Huberman, A. M. (1994). Qualitative Data Analysis*:* An Expanded Sourcebook *(2nd edition). Thousand Oaks CA: Sage Publications.*
Miles, M. B. & Weitzman, E. (1994). "Choosing Computer Programs for Qualitative Data Analysis". In M. B. Miles & A. M. Huberman. Qualitative Data Analysis*:* An Expanded Sourcebook *(2nd edition). Thousand Oaks CA: Sage Publications, pp. 311-317*
Miller, L. (1989). Musical savants*:* exceptional skill and mental retardation*. Hillsdale, New Jersey: Laurence Erlbaum.*
National Health and Medical Research Council (NHMRC), 2007. National Statement on Ethical Conduct in Human Research*. Joint statement of the Australian Research Council (ARC), National Health and Medical Research Council (NHMRC) and Universities Australia (UA). Retrieved from <http://www.nhmrc.gov.au/publications/synopses/r39syn.htm.>.*
Nelson, D., & Cox, M. (2013). Lehninger Principles of Biochemistry *(6th edition). New York: W.H. Freeman.*
Ockelford, A (2007). **In the Key of Genius*:* The Extraordinary Life of Derek Paravicini***. London: Hutchinson.***
Ockelford, A (2008). Music for Children and Young People with Complex Needs*. London: Oxford University Press.*
Ockelford, A (2012a). **Mapping Musical Development in learners With The Most Complex Needs*:* The Sounds of Intent Project***. In G. E. McPherson & G. Welch (editors), Oxford Handbook of Music Education. New York: Oxford University Press.***
Ockelford, A (2012b). **Special Abilities, Special Needs*:* Commentary***. In G. E. McPherson & G. Welch (Eds.), Oxford Handbook of Music Education. New York: Oxford University Press.***

Office of Standards in Education (OFSTED), 2012a. Music in Schools*:* Wider Still, and Wider *(*Quality and Inequality in Music Education 2008-2011*). London: Office of Standards in Education. Retrieved from <http://www.ofsted.gov.uk/publications/110158>.*

Office of Standards in Education (OFSTED), 2012b. Retrieved from <www.youtube.com/ofstednews>.

Pascoe, R., Leong, S., MacCallum, J., Mackinlay, E., Marsh, K., Smith, R (Bob)., Church, T., Winterton, A. (2005). National Review of School Music Education*:* Augmenting the Diminished. *Canberra: Department of Education, Science and Training. Retrieved from <http://www.dest.gov.au/sectors/school_education/publications_resources/profiles/school_music_education.htm>.*

Patton, M. Q. (2002). **Qualitative Research and Evaluation Methods *(**3rd edition**).** *Thousand Oaks CA: Sage Publications.*

Peretz, I & Zattore, R.J (editors) (2003). The Cognitive Neuroscience of Music*. Oxford/New York: Oxford University Press.*

Qualifications and Curriculum Authority (Department for Education and Employment) (2001). Planning, Teaching and Assessing the Curriculum for Pupils with Learning Difficulties*. Accessed from <http://www.nc.uk.net/ld/index.html>.*

Radocy, R. E. & Boyle, J. D. (2003). The Psychological Foundations of Musical Behaviour *(4th edition). Springfield IL: Charles C Thomas.*

Sacks, O (2007). Musicophilia*:* Tales of Music and the Brain. *New York.*

Schapira, A.H.V. (editor) (2007). Neurology and Clinical Neuroscience*. Philadelphia: Elsevier Mosby.*

Spradley, J. P. (1980). Participant Observation*. New York: Holt, Rinehart and Winston.*

Spring, B. & Hitchcock, K. (2009). "Evidence-based practice in psychology". In I.B. Weiner & W.E. Craighead [editors]. Corsini's Encyclopedia of Psychology *(4th edition), pp. 603-607. New York: Wiley.*

Stake, R. E. (1978). "The Case Study Method in Social Inquiry". Educational Researcher*, 7(2), pp. 5-9.*

Stake, R. E. (1983). "The Case Study Method in Social Inquiry". In G. F. Madaus, M. S. Scriven & D. L. Stufflebeam (editors), Evaluation Models*. Boston: Kluwer-NijHoff, pp. 279-286.*

Stake, R. E. (1988). "Case Study Methods in Educational Research: Seeking Sweet Water". In R. M. Jaeger (editor), Complementary Methods for Research in Education*. Washington DC: American Educational Research Association, pp. 253-278.*

Stake, R. E. (1995). The Art of Case Study Research*. Thousand Oaks, CA: Sage Publications.*

Thomas, G & Pring, R [editors] (2004). Evidence-based Practice in Education*. Maidenhead, Berkshire, UK: Open University Press.*

Thompson, W (2009). Music, Thought and Feeling: *Understanding the Psychology of Music. Oxford: Oxford University Press.*

Wanna, J., Butcher, J. & Freyens, B. (2010). Policy in Action*:* The Challenge of Service Delivery*. Sydney: UNSW Press.*

Yin, R. K. (2009). Case Study Research: Design and Methods *(4th edition). Thousand Oaks, CA: Sage Publications.*

Yin, R. K. (2012). Applications of Case Study Research *(3rd edition). Newbury Park, CA: Sage Publications.*

WHEN THEY START THINKING IN SOUND – ASSESSING THE EAR TRAINING SKILLS OF FIRST-YEAR MUSIC EDUCATION STUDENTS

Anna Wolf[1], Reinhard Kopiez[1]
[1]Hanover Music Lab, Hanover University of Music, Drama and Media, Germany
Time: 10:00

1. Background: Institutes of higher music education lack valid instruments to measure the performance level and development of their students (Wolf & Kopiez, 2014). Our team of music psychologists and music theorists developed a collection of items for the assessment of ear training skills. These items cover melodic, harmonic and rhythmic facets of ear training (Karpinski, 2000). We based our selection on Item response theory, more precisely the Rasch model, to best achieve an objective, reliable and valid test construction (e.g. Bond & Fox, 2007). The items have been pre-tested at music schools.

2. Aims: We are in the process of measuring the development in ear training of first-year music education students over a period of eight months.

3. Method: In the current study, we use a subset of these items (N=21) to measure the performance development of a complete cohort of students (N=37, M=20 years). In a pre-post design, regular lessons in music theory and ear training are regarded as intervention.

4. Results: The first measurement showed a suitable test difficulty for the sample of first-year music students (students' response probability M=0.55, SD=0.21, range=0.19-0.90). However, several cautious biases were detected: Men performed better than women (t=1.71, df=24.8, p=.099, d=0.6), and older students performed worse than younger students (Spearman's ρ=-.40, p=.01). The test score correlated positively with the cumulated years of playing harmony instruments (r=.36, df=35, p=.03) and negatively for melody instruments (r=-.19, df=35, p=.26). The second measurement (post-test) will be conducted in the near future.

5. Discussion: This is the first study to investigate the development of music students' ear training skills which uses a standardized and empirically scrutinized instrument. Both the range of response probabilities and item difficulties show a good mix of easy, motivating items as well as hard items for a successful discrimination. In our second measurement, we are going to use old ("anchor") and new items to be able to both compare the test scores and avoid memory effects (Bond & Fox, 2007). While our current results are based on a rather small sample, the upcoming second measurement will review the preliminary results and objectively quantify the students' development.

[9D] Ears and Eyes

LIÈGE 09:00- 10:30 Friday 8 Aug 2014

THE EFFECT OF EXPOSURE TIME ON THE INTERMODAL EFFECT OF AUDIO-VISUAL STIMULUS: COMPARISON BETWEEN EFFECT OF MUSIC AND EFFECT OF PAINTINGS.

Teruo Yamasaki[1]
[1]Osaka Shoin Women's University, Japan
Time: 09:00

Author's previous study found that musical effect on the perceived impression of paintings was larger than visual effect on the perceived impression of music when music and paintings were presented simultaneously. As several researchers reported, music can draw positive or negative emotion very rapidly. Similarly, visual stimulus like facial expression can also draw such emotions with very short exposure time.

However, there is a crucial difference between music and paintings. Music continues to change during listening, whereas paintings are still during presentation. Therefore, music might continue to draw emotion during listening, but paintings might develop perceptual adaptation during their presentation. This difference of the temporal variability between music and paintings might cause the asymmetry of intermodal effect on the perceived impression of music and paintings. In that case, it is expected that to make the exposure time shorter decreases the musical effect, whereas the visual effect is not influenced by changing exposure time. Consequently, when music and paintings are presented with relatively short exposure time, the asymmetry of intermodal effect between music and paintings might become smaller or disappear. This study investigated this possibility. Participants were presented music and paintings individually or simultaneously, and were asked to rate their impression along two dimensions, namely potency and brightness. Music and paintings were presented with 4 or 20 seconds in this study, while those were presented with 80 seconds in the author's previous study. Results revealed that the musical effect on the perceived impression of paintings were significantly larger than the visual effect on the perceived impression of music in both conditions of exposure time. Contrary to the expectation, the exposure time had no effect on the intermodal effect. These results are discussed from a view point of the temporal variability of music and paintings.

COMPARISON OF THE PROCESS FOR CREATING PERCEIVED CONGRUENCE BETWEEN A MOVING PICTURE AND SOUND BASED ON SYNCHRONIZATION OF AUDITORY AND VISUAL ACCENTS AND SIMILARITY OF AUDITORY AND VISUAL AFFECTIVE IMPRESSIONS

Saki Fujiyama[1], Fumiyuki Takishita[2], Toru Yahagi[3], Shin-Ichiro Iwamiya[1]
[1]Faculty of Design, Kyushu University, Japan, [2]Graduate School of Design, Kyushu University, Japan, [3]School of Design, Kyushu University, Japan
Time: 09:30

ABSTRACT

The aim of this study is to compare the psychological process of creating formal congruency and semantic congruency by continuous-rating experiments of congruence between a moving picture and sound. The experimental stimuli were combinations of simple sound and moving picture. The audio-visual stimuli for formal congruency were synchronized and asynchronized auditory and visual patterns. Those for semantic congruency were audiovisual combinations of similar and opposite affective impressions. Experiments were conducted by the continuous rating method. Participants were instructed to rate instantaneously congruence between a moving picture and sound. We compared the durations to reach the saturation point of congruence scores between formal congruency and semantic congruency conditions. The duration to reach the saturation point of congruence based on the synchronization of auditory and visual accents was shorter than that based on the similarity of auditory and visual affective impressions. Furthermore, although synchronization of auditory and visual accents was broken, participants in the experiment tended to find synchronized portions, and the perceived congruence thus rapidly fluctuated. In contrast, the perceived incongruency based on opposite audiovisual affective impression was stable. It takes longer duration to reach the saturation point of semantic congruency than to reach that of formal congruency.

1. INTRODUCTION

Visual media productions, such as movies and television programs, are rarely composed purely of motion pictures. They are a combination of motion pictures and sound. A vital element in making visual media presentations more impressive is "congruence" between moving pictures and sound. Perceived congruence between sound and moving pictures has two aspects (Bolivar, Cohen & Fentress, 1994). One is formal congruency: the matching of auditory and visual temporal structures. The other is semantic congruency: the similarity between auditory and visual affective impressions. Formal congruency provides a united perceptual form to auditory and visual information, whereas semantic congruency helps to communicate the meaning of audiovisual content to the perceiver. Both types of congruency play important roles in the cross-modal interaction of disparate events in auditory and visual domains.

Previous empirical studies showed formal and semantic congruency were actually effective to raise the perceived congruence and evaluation of audiovisual stimuli (Lipscomb, 1996, Iwamiya et al., 2002). However, in the previous studies on congruence between moving pictures and sound, participants were asked to rate the perceived congruence of overall audiovisual stimuli after watching and listening the entire stimuli. These studies have not focused on the psychological process for creating audiovisual congruence. The aim of the present study is to compare the psychological process of creating formal congruency and semantic congruency by continuous-rating experiments of congruence between a moving picture and sound. The difference of both types of audiovisual congruency was discussed by the temporal features of momentary perceived congruence.

2. MEASUREMENT OF OPERATION DURATION OF COMPUTER MOUSE FOR CONTINUOUS MEASUREMENT

In continuous rating experiments, participants rate subjective congruence between a moving picture and sound using the pointer on a rating scale on a display controlled by a computer mouse while watching and listening to audiovisual stimuli. At the beginning of the experiment, the pointer is located on the neutral point of the rating scale. Therefore, the manipulation time might be necessary to move the pointer to the proper position on the scale after AV stimuli were presented. The operation time might affect the interpretation of rating results of continuous measurement. Therefore, as a preliminary experiment, we measured the required time of controlling a computer mouse to shift the pointer from the neutral potion to one of end points of the scale on a computer display.

2.1. Method

The visual stimulus was the transformation from black screen (illuminance level on the display was 3.53 lx) to blue screen (illuminance level on the display was 103 lx) of a computer display. Auditory stimulus was one-shot sound of a bass drum. The auditory stimulus was composed by Cakewalk Music Creator 5 and the visual stimulus was composed by Adobe Premiere 6.0. Stimulus presentation was controlled by a personal computer (Dell OPTIPLEX 745). The auditory stimulus was presented to the participants through the headphones (STAX SR-307) at a level of 60 dB (A-weighted equivalent sound pressure level). The visual stimulus was presented on a computer display (17 inches). The distance from the display to participants' eyes was approximately 0.7 m.

Five students (1 males and 4 females) of Kyushu University participated in this study.

Participants were instructed to move the pointer from the center to the right edge of the rating scale as soon as possible after the auditory stimulus was presented or the display changed to the blue screen from the black screen by operating a computer mouse. The operating duration of computer mouse was the duration from the time of presenting stimulus to the time when the pointer reached the right edge of the rating scale. Stimulus presentation and data acquisition was controlled by the original software developed on Microsoft excel VBA. Position of the pointer on the scale and time of every movement of the pointer was recorded by this software.

2.2. Results

The operating duration of the computer mouse for each participant was 0.77 s~1.16 s (the average value was 0.90 s). According to t test, the difference between the average operating duration for the auditory stimulus and that for the visual stimulus was not statistically significant ($p>.1$). Because the operating duration of the computer mouse was almost within 1 s, this time gap might not seriously affect the interpretation of the experimental results obtained from continuous rating.

3. EXPERIMENT 1: CONTINUOUS RATING OF CONGRUENCE BETWEEN MOVING PICTURE AND SOUND FOR SYNCHRONIZED AND ASYNCHRONIZED AUDITORY AND VISUAL PATTERNS

In the first experiment, we conducted the continuous rating experiment of congruence between moving picture and sound for the audiovisual stimuli whose successive auditory and visual accents were synchronized or not synchronized.

3.1. Method

Fifteen students (13 males and 2 females) of Kyushu University participated in this experiment. The audiovisual stimuli were the combinations of repetitive simple quadruple time pattern of a base drum (120 bpm) and the repetitive transformation of four color images (pink, blue, green and yellow) on a computer display. In the case of "synchronized stimulus," the timing of a beat and that of color change were synchronized: the color of the screen changed when the first and third beat of each bar were played. In the case of "asynchronized stimulus," the timing of a beat and that of color change were not synchronized. Similar to the synchronized stimulus, the rhythmic pattern of auditory stimulus was periodic, however the changing pattern of colors of the screen was not periodic. The duration of these stimuli were 90 s. "Synchro-asynchro-synchro stimulus" was composed of synchronized stimulus (30 s), asynchronized stimulus (30 s) and synchronized stimulus (30 s). "Asynchro-synchro-asynchro stimulus" was composed of asynchronized stimulus (30 s), synchronized stimulus (30 s) and asynchronized stimulus (30 s). The auditory stimuli were presented at a level of approximately 60 dB (A-weighted equivalent sound pressure level).

The task of participants was to rate the level of congruence between moving picture and sound (from -1 to +1) continuously during the presentation of AV stimuli by operating a computer mouse. The starting point of the rating experiment was the center point of the scale (0). A psychophysical measurement software (EMujoy, Nagel et al., 2007) was used to measure the continuous rating value of perceived congruence between moving picture and sound. The positions of the pointer on the scale (-1 ~ 1) and time every movement were recorded by this software. From these records, we detected the scale values of the pointer at every 10 ms as "congruence level" at each moment.

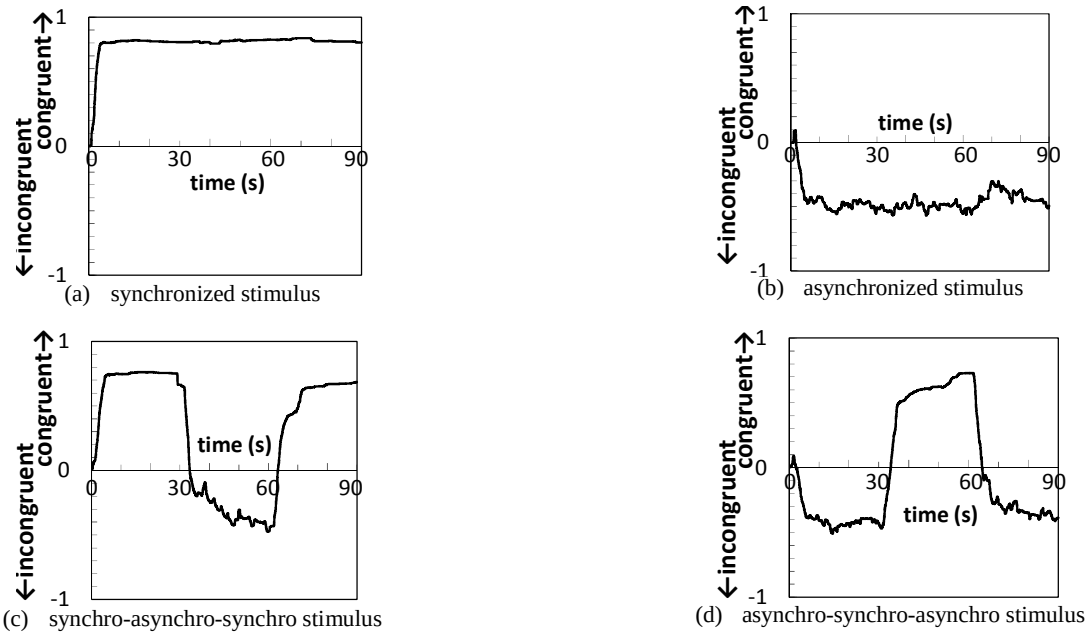


Figure 1: The instantaneous level of perceived congruence of each AV stimulus used in the Experiment 1 (the average values of congruence at each 10 ms moment among all the participants)

3.2. Results and Discussion

Figure 1 shows the instantaneous level of perceived congruence of each AV stimulus used in the first experiment. These levels are the average values of congruence at each (10-millisecond) moment among all the participants. The average value of congruence from 0 to 90 s of "synchronized stimulus" was 0.76 and that of asynchronized stimulus was -0.45. According to t test, the difference of these average values of

congruence between these two stimuli was statistically significant (*p<.001*). When the visual and auditory accents were synchronized, the perceived congruence was higher than when they were not synchronized. This result indicated that the participants perceived congruence between moving picture and sound based on the formal congruency. This result is consistent with the previous study that was conducted the rating experiment after listening and watching the audiovisual stimuli (Lipscomb, 1996). The average values of congruence from 0 to 30 s, 30 to 60 s and 60 to 90 s of the synchro-asynchro-synchro stimulus were 0.69, -0.21 and 0.47, respectively. According to multiple comparison tests (Bonferroni method) among these durations of the synchro-asynchro-synchro stimulus, the difference between the average value from 0 to 30 s and that from 30 to 60 s, and the difference between the average value from 60 to 90 s and that from 30 to 60 s were statistically significant (*p<.001, p<.01, respectively*). Similar tendency was observed in the asynchro-synchro-asynchro stimulus. The participants sensitively perceived the difference of congruence between synchronized and asynchronized condition.

The level of congruence has been increased linearly just after the synchronized condition was presented and reached to the saturation level within 4 s. After a physical condition of formal congruency between moving picture and sound (synchronized pattern) was provided, it might take about 4 s that the perceived congruence reached to the saturation level. Furthermore, once the congruence level reached to this saturation level, the congruence level was constant while the condition of formal congruence was kept. On the other hand, the congruence level of asynchronized conditions was fluctuated even after the saturation point of congruence. In this condition, although the changing pattern of colors on the display was not periodic, the similar periodic rhythmic pattern to the synchronized condition was presented. The participants might tend to find synchronized portions from the successive auditory and visual accents.

4. EXPERIMENT 2: CONTINUOUS RATING CONGRUENCE BETWEEN MOVING PICTURE AND SOUND FOR AUDITORY AND VISUAL PATTERNS OF SIMILAR AND OPPOSITE AFFECTIVE IMPRESSIONS

4.1. Preliminary Experiment to Select Auditory and Visual Materials

In the second experiment, we conducted the continuous rating experiments of congruence between moving picture and sound of the audiovisual stimuli whose visual and auditory affective impressions were similar or opposite. Therefore, selection of visual and auditory materials arousing similar or opposite affective impressions was necessary. We conducted a preliminary experiment to select visual and auditory materials of cheerful or gloomy affective impressions.

Five students (2 males and 3 females) of Kyushu University participated in this study.

The auditory materials were periodic repetitive patterns of a major triad (the tonic is the note C, 261.63 Hz) or a minor triad (the tonic is the note C, 261.63 Hz). The duration of one each triad was approximately 4.4 s and each triad was presented periodically. The total duration of each auditory stimulus was 90 s. Visual materials were three smiling facial expressions and three crying facial expressions excerpted from commercial movies. The duration of each visual stimulus was 90 s. When the original visual material was shorter than 90 s, it was played repeatedly until the duration reached to 90 s. The experimental equipment was same as used for the first experiment.

The task of participants was to rate the level of cheerfulness (from -1 to +1) of the visual or auditory stimuli continuously during the presentation of stimuli by operating a computer mouse. The initial point of the rating experiment was the center point of the scale (0). Stimulus presentation and data acquisition was controlled by the original software (see 2.1).

Positions of the pointer on the scale at every 10 ms were recorded by this software. From these records, we defined the scale values of the pointer at every 10 ms as “cheerfulness level” at each moment.

The cheerfulness level of the major triad stimulus was saturated at 0.4 and that of minor triad was saturated at -0.6. Furthermore, both of the cheerfulness levels were almost constant after reaching the saturation level. We decided to use periodic patterns of major triad and minor triad as the experimental auditory stimuli. The periodic pattern of major triad is called A1, and that of minor triad is called A2.

The each cheerfulness level of the three moving pictures of smiling facial expressions was saturated almost at 0.6 and that of the three moving pictures of crying facial expressions was also almost at -0.6. They could be used as visual stimuli, however the visual material was also required not to change their affective impression during its presentation, because the creating process based on semantic congruency would be discussed in the second experiment. Therefore, we selected one smiling and one crying facial expression of which the cheerfulness level was the most stable. The selected visual material of smiling facial expression was called V1, and that of crying facial expression is called V2.

4.2. Continuous Rating of Congruence between Moving Picture and Sound

Method The audiovisual stimuli were combinations of visual and auditory materials selected in the preliminary experiment.

The audiovisual stimulus S1 consisted of the visual stimulus V1 and the auditory stimulus A1, and both A1 and V1 aroused cheerful affective impression. In contrast, the audiovisual stimulus S2 consisted of the visual stimulus V2 and the auditory stimulus A2, and both A1 and V1 aroused gloomy affective impression. The affective impressions of visual and auditory stimuli of S1 and S2 were similar and expected to create semantic congruency. The audiovisual stimuli D1 and D2 were combinations of moving picture and sound of opposite affective impression. D1 was the combination of V2 and A1, and D2 was the combination of V1 and A2. D1 and D2 might not create semantic congruency. The duration of these audiovisual stimuli was 90 s.

Furthermore, the combinations of S1 and D1, those of S2 and D2 were used as the audiovisual stimuli. In these combinations, the duration of each part of the stimuli was 30 s. The stimulus S1-D1-S1 was composed of S1 (30 s), D1 (30 s) and S1 (30 s). D1-S1-D1 was composed of D1 (30 s), S1 (30 s) and D1 (30 s). S2-D2-S2 was composed of S2 (30 s), D2 (30 s) and S2 (30 s). D2-S2-D2 was composed of D2 (30 s), S2 (30 s) and D2 (30 s).

The experimental equipments were the same as used in the first experiment. Similar to the first experiment, the task of participants was to rate the level of congruence between moving picture and sound (from -1 to +1) continuously during the presentation of AV stimuli by operating a computer mouse. Stimulus presentation and data acquisition was controlled by the original software (see 2.1). Positions of the pointer on the scale

at every 10 ms were recorded by this software. From these records, we defined the scale values of the pointer at every 10 ms as "congruence level" at each moment.

Sixteen students (10 males and 6 females) of Kyushu University participated in this study.

Results and discussion Figure 2 shows the instantaneous level of perceived congruence of each AV stimulus used in the second experiment. These levels were the average values of congruence among all the participants at the interval of each 10 ms.

The average value of congruence from 0 to 90 s of the stimulus S1 was 0.44 and that of S2 was 0.55. The average value of congruence from 0 to 90 s of the stimulus D1 was -0.28 and that of D2 was -0.36.

According to the multiple comparison tests (Bonferroni method), the differences of average values of congruence between S1 and D1, those between S1 and D2, those between S2 and D1, and those between S2 and D2 were statistically significant respectively (*p<.001 respectively*). When the auditory and visual affective impressions were same, the perceived congruence was higher. In contrast, when they were opposite, the perceived congruence was lower. This tendency indicated that the participants perceived congruence between moving picture and sound based on the semantic congruency. This tendency is consistent with the previous study conducted the rating experiment after listening and watching the audiovisual stimuli (Iwamiya et al., 2002).

The average value of congruence from 0 to 30 s, that from 30 to 60 s and that from 60 to 90 s of S1- D1- S1 were 0.44, -0.30 and 0.30, respectively. According to multiple comparison tests (Bonferroni method) , the difference between the average value from 0 to 30 s and that from 30 to 60 s, and the difference between the average value from 60 to 90 s and that from 30 to 60 s were statistically significant (*p<.001, p<.01, respectively*). Similar tendency was observed from S2- D2- S2, D1- S1- D1, and D2- S2 - D2. The congruence level was higher when the semantic congruency was created than when it was not created. When the status of the semantic congruency changed, the participants sensitively perceived the difference of congruence.

The level of congruence have been increased gradually after the S1 and S2 were presented and reached to the saturation level within 10 to 15 s. Compared to formal congruency, it takes more time to reach to the saturation level of congruence in the case of semantic congruency. The semantic congruency between visual and auditory information was created after visual and auditory affective impressions were perceived and their impressions were compared. The level of congruence based on semantic congruency might have been increased moderately, because the process to create congruence between auditory and visual information was more complex and longer duration was necessary than in the case of formal congruency.

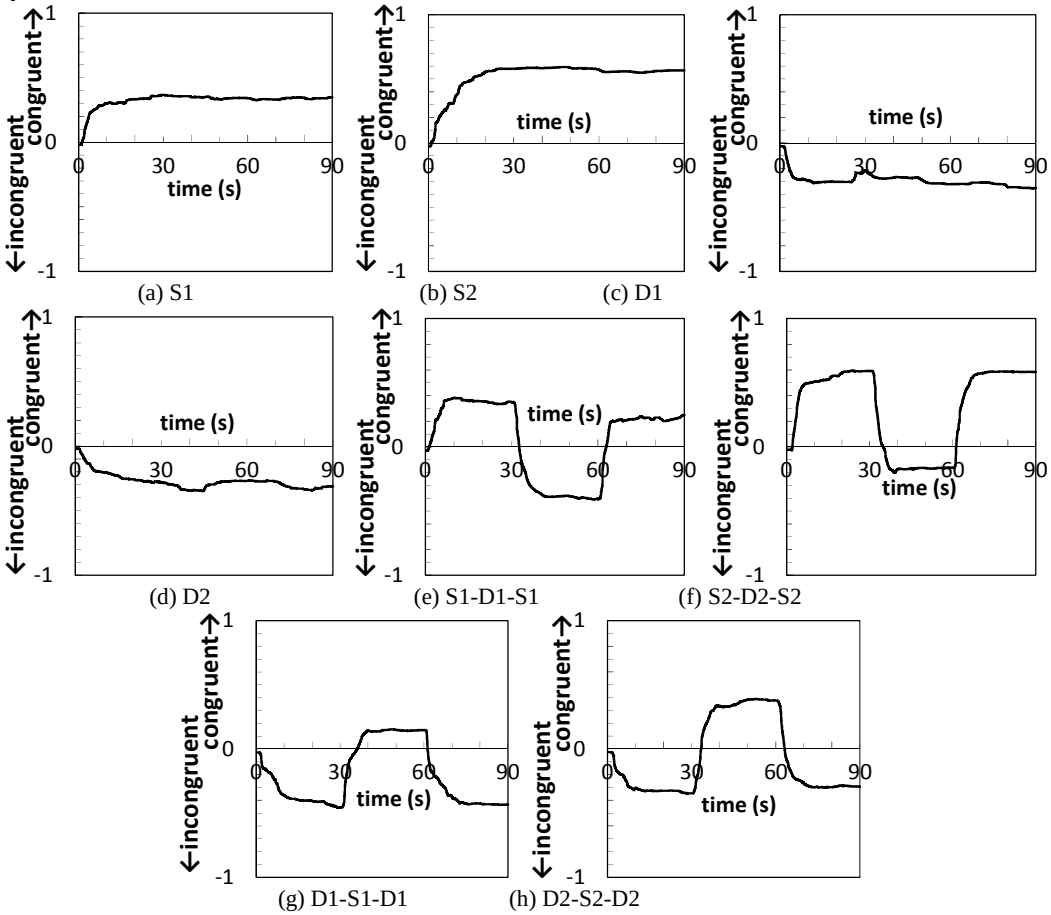


Figure 2: The instantaneous level of perceived congruence of each AV stimulus used in the Experiment 2 (the average values of congruence at each 10 ms moment among all the participants)

On the other hand, the congruence level of D1 and D2 have been decreased and reached to the saturation level of incongruence. In these conditions, the auditory and visual affective impressions were opposite, so that the semantic congruency was not created. Therefore, similar to asynchronized conditions, the congruence level was low. However, in contrast to asynchronized conditions, the perceived incongruency based

on opposite audiovisual affective impression was stable. The participants might not to seek similarity between visual and auditory affective impression when the affective impressions were opposite.

5. COMPARISON OF THE PROCESSING BETWEEN FORMAL CONGRUENCY AND SEMANTIC CONGRUENCY

According to the experimental results shown in Figures 1 and 2, the duration to reach saturation point of perceived congruence based on formal congruency seemed to be shorter than that based on semantic congruency. The difference of duration to reach saturation level of congruence might reflect the difference of creating process of congruence between moving picture and sound. In this chapter, the difference of duration to reach saturation level of perceived congruence among the conditions was discussed quantitatively.

In each participant's data, the level of congruence between moving picture and sound increased monotonically from the stimulus presentation to the saturation point within 30 s in any condition. To compare the durations to reach saturation point of perceived congruence among the experimental conditions quantitatively, we defined the duration to reach the saturation point of perceived congruence (DSC) as the duration from the stimulus presentation to reach 95% of the congruence value at 30 s. Then, we compared the value of DSC in the case of formal congruency with that of two types of semantic congruency: cheerful congruency and gloomy congruency. We calculated the DSC in the case of formal congruency from the data synchronized stimulus and synchro-asynchro-synchro stimulus; In both the stimuli, the participants started rating from the synchronized condition. Similarly, we calculated the DSC in the case of semantic congruency of cheerful affective impression from S1 and S1-D1-S1 that started from the combination the periodic pattern of major triad and smiling facial expression, and that of gloomy affective impression from S2 and S2-D2-S2 that started from the combination the periodic pattern of minor triad and crying facial expression. Although this procedure was arbitrary, we could compare the duration to reach saturation point of perceived congruence between formal congruency and semantic congruency quantitatively with the same criterion.

Table 1: The average values of DSC (the duration to reach the saturation point of perceived congruence) among the participants in formal congruency and those in two types of semantic congruency

	formal	cheerful semantic	gloomy semantic
DSC (s)	3.8	9.1	12.0
SD	1.6	8.9	8.3

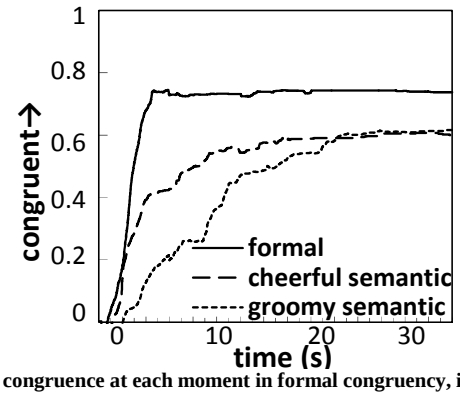


Figure 3: The instantaneous level of perceived congruence at each moment in formal congruency, in cheerful semantic congruency, and in gloomy semantic congruency

Table 1 shows the average values of DSC among the participants in formal congruency and in two types of semantic congruency. Figure 3 shows the average values of instantaneous level of perceived congruence at each moment in formal congruency, in cheerful semantic congruency, and in gloomy semantic congruency, respectively: the creating process of congruence between moving picture and sound in each case. In the case of formal congruency, the average value of DSC was 3.8 s. The average value of DSC in the case of cheerful semantic congruency was 9.1s, and that in the case of gloomy semantic congruency was 12.0 s.

The average values of DSC in both types of semantic congruency were longer than the duration for the auditory or visual elements to reach the saturation level of the cheerful or gloomy affective impression (A1: 8.5 s, A2:8.4 s, V1:8.5 s, V2:6.6 s). The duration to reach the saturation point of perceived congruence was longer than that of affective visual and auditory impressions.

The average values of DSC in the cases of cheerful and gloomy semantic congruency were longer than that in the case of formal congruency. Kruskal-Wallis test revealed the effect of the conditions on the average values was statistically significant ($p<.001$). According to the Mann-Whitney U test using Ryan's method, the difference between the average value of DSC in formal congruency and that in cheerful semantic congruency was statistically significant ($p<.05$). The difference between the average value of DSC in formal congruency and that for gloomy semantic congruency was marginally significant ($p<.1$). The difference between the average value of DSC in cheerful semantic congruency and that for gloomy semantic congruency was not statistically significant ($p>.1$).

The quantitative analysis confirmed that the duration to reach the saturation point of perceived congruence based on formal congruency was shorter than that based on semantic congruency. This difference reflected the difference of the process for creating perceived congruence between moving picture and sounds in both types of congruency. Formal congruency was created just by the detection of synchronization between auditory and visual accents. Therefore, formal congruency was created quickly. On the other hand, semantic congruency was created by the matching of visual and auditory affective impressions. Before judging the matching of auditory and visual affective impression, the perceiving process of affective visual and auditory impression was necessary. The difference of creating time of perceived congruence revealed that the psychological process for creating semantic congruency was more complex than that for formal congruency. Therefore, it took more time for creating semantic congruency than for formal congruency.

6. CONCLUSION

The creating process of formal congruency and semantic congruency was measured by continuous-rating experiments of congruence between a moving picture and sound. The duration to reach the saturation point of semantic congruency was longer than that of formal congruency. This tendency revealed that the psychological process for creating semantic congruency was more complex than that for formal congruency. When the condition of formal congruency was broken, the participants tended to find portions of formal congruency. In contrast, they did not when the condition of semantic congruency was broken.

7. ACKNOWLEDGMENTS

This work was supported by Grant-in-Aid for Challenging Exploratory Research No. 25560006 from MEXT Japan.

8. REFERENCES

Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effect on the interpretation of visual action. Psychomusicology, 13, *28-59.*

Iwamiya, S., Jogetsu, U., Sugano, Y. & Takada, M. (2002).The effects of musical tempo and tonality and visual speed and density on perception of audiovisual content. Jornal of Music Perception and Cognition, 8, *53-64 (In Japanese).*

Lipscomb, S. D. (1996). Cognition of musical and visual accent structure alignment in film and animation. Proceedings of 4th International Conference of Music Perception and Cognition, *309-313.*

Nagel, F., Kopies, R., Grewe, O., & Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. Behavior Research Methods, 39(2), *283-290.*

THE INFLUENCE OF TONALITY DURING SIGHT-READING PERFORMANCE

Olivia M. Podolak[1], Mark A. Schmuckler[1]
[1]University of Toronto Scarborough, Canada

Time: 10:00

Psychological research over the past decades has demonstrated that the tonal hierarchy influences on-line processing of musical information through the creation of expectancies. The present study extends this work by investigating the role that the tonal hierarchy plays during sight-reading performance. To do this, we compared the sight-reading accuracy of twelve experienced pianists across three tonal constructs (major, minor and atonal) and three musical textures (monophonic, homophonic and polyphonic). It was hypothesized that sight-reading performance would be the worst in instances with no tonal information, as the participants would be unable to generate appropriate expectancies to guide their sight-reading. The results indicated that pianists performed the major excerpts with significantly greater accuracy than the atonal excerpts only. Furthermore, the errors made within the major excerpts were significantly more diatonic than non-diatonic, clearly demonstrating a bias towards a tonal construct as opposed to an atonal one. The existence of this bias is supported by the fact that a significant portion of the atonal excerpts was actually performed tonally by the participants. Interestingly, the diatonic bias was not found in the minor excerpts, suggesting that the minor hierarchy does exert as much of an influence during sight-reading as the major tonality. This is in line with previous research, which has indicated that the minor tonality is less psychologically stable than the major tonality. Future directions include investigations into performance tempo and how it might interact with tonality during sight-reading, as well as applications of mobile eye-tracking technology to the present paradigm.

Session 10

[10A] Music and Memory 3

SEOUL 11:00-12:30 Friday 8 Aug 2014

THE CONTOURS OF CATCHINESS, OR WHERE TO LOOK FOR A HOOK

John Ashley Burgoyne[1], Jan Van Balen[2], Dimitrios Bountouridis[2], Themistoklis Karavellas[1], Frans Wiering[2], Remco C. Veltkamp[2], Henkjan Honing[1]
[1]Universiteit Van Amsterdam, Netherlands, [2]Universiteit Utrecht, Netherlands

Time: 11:00

1. BACKGROUND: Given a piece of music, where should one look for the 'hook'? Conventional wisdom would say the chorus or refrain, but the story in the scientific literature is more nuanced. Fundamentally, hooks are triggers for recalling musical imagery to memory, and there is tension between theories that hooks' triggering power comes from repetition and theories that this power instead arises from sudden, striking conflicts between static, enculturated expectations from a piece of music (schematic expectation) and dynamically-updated expectations during particular listenings (adaptive expectation). At SMPC 2013, we presented a novel experimental design for studying hooks with a music recognition game for smartphones. We have since executed our first experiment with this design, and this presentation will highlight the key results.

2. AIMS: Our analysis considers the following hypotheses. 1. There are two primary 'catchiness contours' in pop music: (a) the ONE-MORE-TIME CONTOUR, in which the easiest-to-recall fragment returns one or more times, alternating with less memorable fragments, and (b) the ONE-TIME-ONLY CONTOUR, in which a single, striking moment dominates all other fragments. 2. Pieces of music with a one-time-only contour are more memorable overall than pieces of music relying on a one-more-time contour.

3. METHOD: Working with 14 000 song fragments, corresponding to the first 15 s of the major structural sections (verse, chorus, bridge, etc.) of 1600 popular songs, we used the game to collect response times for how quickly participants were able to recognise or to be confident that they did not recognise each fragment. Ratcliff's drift-diffusion model (1978) can incorporate both positive and negative responses into a single statistical analysis.

4. RESULTS: The experiment attracted 1600 participants and yielded 130 000 responses. As expected, there were significant differences in response times across different fragments ($p < .001$). We will present drift-diffusion models fit with the HDDM toolbox (Wiecki, Sofer & Frank, 2013) to explain these differences and by which cognitive mechanisms they relate to our principal hypotheses.

5. CONCLUSIONS: This experiment contributes to an important debate in the literature on the relationship between repetition and (violated) expectation and will be a critical foundation for future work on specific musical characteristics that stimulate long-term musical memory.

EXPECTING THE CLASSICAL CADENCE: PRIMING EFFECTS

David Sears[1], William E. Caplin[2], Stephen Mcadams[1]
[1]Centre for Research in Music Media and Technology (cirmmt), Schulich School of Music, McGill University. Montreal, Qc, Canada,
[2]Department of Music Research, Schulich School of Music, McGill University. Montréal, Qc, Canada
Time: 11:30

1. Background: Music theorists have often accounted for the perception of closure in tonal music by appealing to theories of expectation. Empirical evidence for musical expectations derives principally either from continuation and expectancy ratings tasks, or from melodic and harmonic priming paradigms, but no study to date has extended these findings to the wide variety of cadential categories found in the "common practice" period. Indeed, Leonard Meyer cited the classical cadence as the quintessential compositional device for suppressing further expectations, as the progression preceding cadential arrival elicits very specific expectations concerning the melodic scale degree, the harmony, and the metric position of the goal event.

2. Aims: The present study extends the current findings in the priming literature to include cadences from the classical style in order to examine how the fulfillment or violation of expectations contributes to the perception of cadential closure.

3. Method: Thirty participants (15 musicians) heard 40 excerpts drawn from Mozart's keyboard sonatas that contained an equal number of perfect authentic (PAC), imperfect authentic (IAC), half (HC), deceptive (DC), and evaded cadences (EV). In each trial, a playback bar was presented on the computer screen to mark the final chord of the excerpt (i.e., cadential arrival). When the cursor reached the final chord, participants were asked to indicate as quickly as possible whether the chord was in or out of tune (in the out-of-tune condition, the entire chord was tuned 40 cents sharp).

4. Results & Conclusions: Linear mixed effects models including fixed effects of cadence category and musical training and crossed random effects of subject and musical stimulus indicated significant effects of cadence category and musical training for response accuracies and reaction times (RTs), with the genuine cadence categories (PAC, IAC, HC) eliciting faster and more accurate responses than the failed categories (DC, EV). The RTs also demonstrated a descending cadential hierarchy that conformed to theoretical models of cadential strength, suggesting that the generation of harmonic and melodic expectations for the moment of cadential arrival may account for the perception of cadential closure during music listening.

AN INFORMATION-THEORETIC APPROACH TO LEARNING, MEMORY, AND EXPECTATION IN UNFAMILIAR MUSICAL STYLES

Kat Agres[1], Marcus Pearce[1]
[1]Queen Mary, University of London, United Kingdom
Time: 12:00

1. Background: Information Theory is useful for measuring listeners' expectations during music listening. In particular, Information Content (IC; negative log probability of an event given context) is a good proxy for psychological unexpectedness. We know little, however, about whether this holds true with increased exposure to music. Also, few studies have examined how variation in Information and Complexity across an entire melody affects momentary expectation for individual tones and memory for the melody.

2. Aims: These studies examined the process of forming and updating musical expectations and recognition memory performance using both computer-generated and ecological melodies. We tested whether measures of Complexity and Information are useful for characterizing both momentary expectation and memory for novel melodies.

3. Method: In our computational approach, tone sequences were created to vary along several information-theoretic measures, to test the impact of this variation on expectation and memory. Expectedness ratings of probe tones were collected in three listening sessions, and each listening session was followed by a recognition memory test. The effect of Information was analyzed per probe tone and per sequence. In a complementary approach, Turkish makam melodies were used to examine the impact of IC and stimulus Complexity on memory for non-Western tonal music. Type of exposure was manipulated, with some participants repeatedly exposed to one set of melodies (Veridical Repetition condition), and others exposed to different sets (Stylistic Familiarity condition).

4. Results: IC values significantly predicted Expectedness ratings for both artificial tone sequences and makam melodies. Interestingly, the average Information of sequences increasingly impacted probe tone Expectedness ratings (in addition to probe tone IC). In the makam study, probe tone IC statistically interacted with melodic Complexity and type of exposure: Low-IC tones were judged more expected when listeners were stylistically familiar with the music than when repetitions of the same stimuli were heard.

5. Conclusions: Our combined computational and ecological approaches elucidate the dynamic process of forming expectations during music listening. We demonstrate that IC accurately captures expectation for non-Western music. In addition, Information and Complexity had varying influences on memory, depending on the type of musical exposure. Lastly, the average Information of sequences affected momentary expectation.

[10B] Preference and Analysis of Sound

KYOTO 11:00-12:30 Friday 8 Aug 2014

PREDICTING MUSIC LISTENING PREFERENCE FROM IMPRESSIONS OF MUSIC PIECES UNDER FOUR MOODS

Daisuke Ishikawa[1], Akira Nishimura[1], Nobuo Koizumi[1]

[1]*Tokyo University of Information Sciences, Japan*
Time: 11:00

ABSTRACT

Our previous research attempted to predict the degree of listening preference toward a piece of music for healing, based on the listener's impression of the music. Multiple-regression analysis showed that the average listening preference scores can be predicted from the impression of the music piece, especially for pieces with a "graceful" impression. Naturally, music listening preferences depend on the listener's mood at that particular moment. The aim of the present study was to predict the degree of music listening preference under four moods (Cheerful, Sad, Relaxed, and Irritated) according to descriptors of musical impression. Twenty musical pieces from various genres were selected from the RWC Music Genre Database. Of the 20 pieces, 17 are found exclusively in the database. In a listening experiment, 51 participants used 22 descriptors of musical impression to evaluate each piece of music on a semantic differential scale. Collected data were used in factor analysis. As a result, the 22 impressions were integrated into five factors: Worth, Brightness, Grace, Brilliance, and Power. The accuracy of the prediction was compared using two techniques: artificial neural network prediction and multiple-regression analysis. Input data for the two techniques were the five factor scores obtained from combinations of the listeners and musical pieces. Correlation coefficients between predicted and estimated listening preferences were calculated for each mood. The results revealed higher correlation under the Cheerful and Relaxed moods than under the Sad and Irritated moods. Correlation coefficients for mean listening preferences obtained from multiple-regression analysis were significantly higher than those obtained from neural network prediction under the Sad and Irritated moods. No significant difference was found between multiple-regression analysis and neural network prediction for the other moods. The average listening preference scores among listeners under the Cheerful and Relaxed moods could be roughly predicted from the average impressions among listeners. However, prediction was less accurate under the Sad and Irritated moods. Multiple-regression analysis showed that both the Worth and Brightness factors were significant for prediction in the Cheerful mood. The Grace factor and Brilliance factor were significant for prediction under the Relaxed mood. In other words, listeners under the Cheerful mood preferred to listen to pleasant and bright music, and listeners under the Relaxed mood preferred to listen to clean and soft music. The correlation coefficient of listening preferences for each listener was lower than that of the mean listening preferences among all listeners.

1. INTRODUCION

Our previous studies attempted to predict the degree of listening preference toward a piece of music for healing, based on the listener's impression of the music (Takagi et al., 2009). Multiple-regression analysis showed that the average listening preference scores can be predicted from impressions of the music piece, especially from pieces with a "graceful" impression.

Naturally, listening preference in music depends on the listener's mood at that particular moment. Therefore, the present study aimed to predict the degree of listening preference in music under four moods (Cheerful, Sad, Relaxed, and Irritated) according to descriptors of musical impression. The accuracy of the prediction was compared using two techniques: multiple-regression analysis and artificial neural network prediction.

Table 1: List of music pieces

Group A		Group B	
RWC Number	Title	RWC Number	Title
2	Forget About It	3	In Your Arms
10	Lost in My Dreams	4	Hold On
19	Asia 2	15	Rocket in My Pocket
29	Bubble	35	Gypsy Eyes
46	Tango Noir	37	Musica Nova
50	<Egmont> Overture, op.84	41	Dance to the Samba
63	`Clair de Lune' from <Suite bergamasque>	69	What Should I Tell Them?
67	Dream Angel	71	Desperate Little Man
87	Je Te Veux	77	Kasheki
94	Sado Okesa	91	Tairyo Bune

2. EXPERIMENT

2.1. Musical Pieces

Twenty musical pieces from various genres were selected from the RWC Music Genre Database. Of these 20 pieces 17 are found exclusively in the database, and were therefore novel to the listeners. From each piece, a 30 s portion was selected to present the impressive part. The pieces were divided into two groups, Group A and Group B, each containing 10 music pieces (Table 1). No. 94 is the only piece that has Japanese lyrics, so that the content of lyrics does not affect the results.

2.2. Method

Participants were 51 young adults aged 19–22 years, of whom 22 were assigned to Group A and 29 were assigned to Group B. The 10 musical pieces for each group were presented in random order. Each piece of music was repeated three times. The participants evaluated musical impression on 22 bipolar, 7--point semantic differential (SD) scales (Iwamiya, 1994) and assessed their degree of listening preference under four moods. The root mean square sound pressure level was set between 55 and 65 dB (A-weighted).

3. ANALYSIS

3.1. Factor Analysis

Scoring for each bipolar item of the SD scale was –3 to +3. Analysis was done by the principal factor method, and varimax rotation was applied to the obtained factor loadings. Five factors were obtained from the factor analysis of the 22 descriptors. The Kaiser–Meyer–Olkin measure of sampling adequacy was 0.864.

Each of the five factors is named to reflect the SD scale items that it includes. The SD scales and their factor loadings are shown in Table 2.

3.2. Predicting listening preference

3.2.1. Prediction method

Multiple-regression analyses and an artificial neural network were used to predict the degree of listening preference under the four moods on the basis of the five factor scores of the music. Correlation coefficients between the evaluated degree of listening preference and the predicted value were calculated as an index of prediction.

Table 2: Semantic differential (SD) scales and their factor loadings

Impression description			Factor Name				
-		+	Power	Worth	Brightness	Grace	Brilliance
Powerless	-	Powerful	0.710	0.292	0.100	0.125	0.209
Weak	-	Strong	0.685	0.153	0.125	-0.040	0.317
Humble	-	Showy	0.650	0.320	0.271	-0.111	0.104
Excited	-	Calm	-0.578	-0.050	-0.054	0.157	-0.138
Poor	-	Rich	0.423	0.389	0.243	0.244	0.022
Present	-	Absent	0.405	0.221	-0.093	0.376	0.110
Unimpressive	-	Impressive	0.226	0.689	0.023	0.213	0.140
Unpleasant	-	Pleasant	0.278	0.667	0.305	0.225	0.099
Interesting	-	Uninteresting	-0.223	-0.613	-0.248	-0.215	-0.106
Common	-	Unique	0.045	0.403	0.046	-0.095	-0.097
Cheerful	-	Gloomy	-0.164	-0.170	-0.814	0.033	-0.025
Dark	-	Bright	0.276	0.254	0.728	0.064	0.118
Light	-	Heavy	0.038	-0.046	-0.687	-0.032	0.104
Dirty	-	Clean	-0.127	0.099	0.003	0.764	-0.074
Neat	-	Mixed	0.267	0.175	0.236	-0.623	-0.051
Vivid	-	Obscure	-0.125	-0.169	-0.219	-0.570	-0.101
Extended	-	Narrow	-0.264	-0.282	-0.117	-0.444	0.081
Cloudy	-	Clear	0.197	0.181	0.214	0.236	0.537
Hard	-	Soft	-0.042	0.068	0.244	0.345	-0.515
Tight	-	Loose	-0.314	-0.053	0.143	-0.094	-0.473
Dull	-	Sharp	0.240	0.360	0.180	0.240	0.423
Realistic	-	Fantastic	-0.122	0.135	-0.034	0.179	-0.271
Percent of variance (%)			27.6	11.9	9.9	6.1	5.0

Table 3: Comparison of correlation coefficients

Target Listening Preference	Method	Mood			
		Cheerful	**Sad**	**Relaxed**	**Irritated**
Individual listener	Multiple-regression analysis	0.670	0.261	0.430	0.095
	Neural network	0.662	0.238	0.402	0.102
Mean among listeners	Multiple-regression analysis	0.906	0.679 **	0.778	0.444 **
	Neural network	0.884	0.560	0.784	0.247

**: significant difference at 1% level

Multiple-regression analyses were performed by using the five factor scores as independent variables and the degree of listening preference as a dependent variable. These analyses were used to predict two types of listening preference: listening preferences of individual listeners and the mean listening preferences among the listeners for each musical piece. To predict individual listening preference, a model was created from the factor scores extracted for 50 of the 51 listeners. The model predicts the degree of listening preference on the basis of the factor scores of the listener for the ten musical pieces in the group's set of pieces. This operation was performed for the 51 listeners. To predict mean listening preference, a model was created from the aggregated factor scores of the 51 listeners. The model predicts the mean degree of listening preference for each musical piece from the mean factor scores of all listeners. This operation was performed for the full set of 20 musical pieces.

Table 4 : Results of multiple-regression analysis

		Cheerful	Sad	Relaxed	Irritated
Multiple Correlation Coefficient		0.686	0.326	0.467	0.199
Multiple regression coefficient	**Power**	0.423 **	0.186 *	-0.081	0.134
	Worth	0.823 **	0.262 **	0.254 **	0.209 *
	Brightness	0.837 **	-0.418 **	-0.001	-0.192 *
	Grace	0.219 **	0.327 **	0.826 **	0.181 *
	Brilliance	0.062	-0.079	-0.352 **	0.033

*: significant at 5% level; **: significant at 1% level

A neural network was used to predict two types of listening preference: listening preferences of the individual listeners and the mean listening preferences among the listeners. Training of the neural network was performed by using the five factor scores as learning data and the degree of listening preference as target data. The neural network adopted the gradient descent back-propagation learning algorithm and were trained for 1000 epochs. To predict individual listening preference, the neural network selected 340 sets of the five factor scores and listening preferences at random from among the 510 sets; the selected sets were used as training data. Following this, listening preferences were predicted for the 170 sets of the unselected five factor scores. To predict mean listening preferences for the 20 musical pieces, the trained neural network used the five mean factor scores from listeners as input data. Correlation coefficients between predicted and estimated listening preferences were calculated after the training and prediction. The above training and prediction cycle in the neural network was repeated 100 times as individual trials. The mean value of the correlation coefficients across the 100 trials is an index representing the prediction accuracy of listening preferences under each mood.

3.2.2. Results of prediction

The correlation coefficients obtained from the multiple-regression analyses and the neural network are compared in Table 3.

Statistical tests on the difference in correlation coefficients between the multiple-regression analyses and the neural network were performed. No significant difference was found in correlation coefficients for individual listeners. Correlation coefficients for mean listening preferences obtained by multiple-regression analysis are significantly higher than those obtained by the neural network for the Sad and Irritated moods. No significant difference was found between results from multiple-regression analysis and those from the neural network in other moods.

The averaged listening preferences among listeners under the Cheerful and Relaxed moods can be roughly predicted from the average impressions among listeners. However, prediction was less accurate under the Sad and Irritated moods.

4. RELATIONSHIP BETWEEN MOODS AND LISTENING PREFERENCE

The results of the multiple-regression analyses are shown in Table 4. Higher coefficients of multiple correlation were found under the Cheerful and Relaxed moods than under the Sad and Irritated moods.

Multiple-regression analysis shows that both Worth and Brightness were significant factors for prediction under the Cheerful mood. Grace and Brilliance were significant factors for prediction under the Relaxed mood. In other words, listeners under the Cheerful mood preferred to listen to pleasant and bright music, and listeners under the Relaxed mood preferred to listen to clean and soft music.

5. DISCUSSION

The reason that the mean listening preferences were better predicted than individual preferences is that the mean factor scores among the listeners are located near the mode of their distribution. As a result, the predicted listening preferences calculated from these mean factor scores are located near the mode of the distribution of the predicted preferences calculated from individual listeners. This causes low mean square errors between the regression line and the predicted mean listening preferences.

An artificial neural network is a computational model that is capable of machine learning as well as pattern recognition. Therefore, neural networks may be more accurate at predicting individual listeners. However, the prediction accuracy of a neural network was similar to that of multiple-regression prediction. The degree of listening preference can be predicted from elements other than impressions of the music. For example, sex, personality, music experience, frequency of listening, and hobbies may represent strong individual differences in listening preferences. These elements, along with the impression of music, can be used as training data for the neural network to predict individual differences in listening preferences.

6. CONCLUSION

Results from the two prediction techniques revealed a higher correlation between prediction and estimation under the Cheerful and Relaxed moods than under the Sad and Irritated moods. Correlation coefficients for mean listening preferences obtained from multiple-regression analysis are significantly higher than those obtained from the neural network under the Sad and Irritated moods. No significant difference was found between multiple-regression analysis and the neural network predictions in other moods. The averaged listening preferences among listeners under the Cheerful and Relaxed moods can be roughly predicted from the average impressions among listeners. However, prediction was less

accurate under the Sad and Irritated moods. Multiple-regression analysis showed that both Worth and Brightness were significant factors for prediction in the Cheerful mood. Grace and Brilliance were significant factors for prediction under the Relaxed mood. In other words, listeners under the Cheerful mood preferred to listen to pleasant and bright music, and listeners under the Relaxed mood preferred to listen to clean and soft music. The correlation coefficients of listening preferences of individual listeners were lower than those of the mean listening preferences among listeners.

7. REFERENCES

Iwamiya. S. (1994). Interactions between auditory and visual processing when listening to music in an audiovisual context: 1. Matching 2. Audio quality. Psychomusicology: A Journal of Research in Music Cognition, 13, 133-153.

Takagi. S., Nishimura. A., & Koizumi. N., (2009) Prediction of favorite music for healing from impression of music using multiple regression analysis. IEICE Technical Report, 109(100), 61-66.

AUDITORY GRAMMAR IN MUSIC

Yoshitaka Nakajima[1], Takayuki Sasaki[2], Kazuo Ueda[1], Gerard B. Remijn[1]
[1]Dept. Human Science, Kyushu University, Japan, [2]Dept. Psychological & Behavioral Science, Miyagi Gakuin Women's University, Japan
Time: 11:30

ABSTRACT

The theoretical framework of Auditory Grammar, in which auditory subevents, i.e., onsets, offsets, fillings, and silences, are concatenated linearly following simple grammatical rules, helps us to understand how we perceive auditory streams, especially melodies in music. Only three types of auditory events are allowed in this grammar: 1) an onset to be followed by a silence, 2) an onset, a filling, and an offset in this order to be followed by a silence, and 3) an onset and a filling in this order to be followed by another auditory event. New auditory demonstrations were created to indicate that these three types of auditory events appear as notes in melodies, but as results of auditory illusions in situations in which the auditory system has to construct melodies connecting auditory subevents belonging to physically separate sounds. These demonstrations show that active commitment of the auditory system is important in music perception.

1. INTRODUCTION

Auditory events and auditory streams are often considered basic units of auditory organization. In many cases, but not always, an auditory event corresponds to a note, and an auditory stream to a melody in music. An auditory stream is a concatenation of auditory events and silences, and we proposed a theoretical framework in which a simple grammar works to construct auditory streams (Nakajima, Sasaki, Ueda, & Remijn, 2014). Such situations correspond to the notation system of Western music.

In our auditory grammar (AG), we have postulated that auditory events are made up of four types of elements: onsets, offsets, fillings, and silences. We try to describe how AG works for music materials. Most auditory events in our everyday life seem to take one of the three modes indicated below: 1) an onset followed by a silence (e.g., a footstep), 2) an onset, a filling, and an offset in this order, followed by a silence (e.g., a whistle), and 3) an onset and a filling in this order, followed by the onset of another auditory event (e.g., a note in a melody played legato). Based on this, we can define an auditory stream: 1) An auditory stream is a linear string of auditory events and silences in time, 2) an auditory stream begins with an onset and ends with a silence, and 3) a silence is not followed immediately by another silence.

This simple set of rules can be used to explain, for example,

1) why a grace note can be a part of the onset of a tone—because two or more onsets very close to each other in time cannot be interpreted as a sequence of onsets grammatically, and are interpreted as a single onset,
2) why a trill can be a substitute of a sustained tone—because a fast succession of onsets cannot be interpreted as such, and is interpreted as an onset, which takes a short period, followed by a filling,
3) why a melody can be heard in a very reverberant church—because an auditory stream consisting of alternating onsets and fillings is extracted as a grammatical and concise interpretation, or
4) why an illusory melody can be heard in a sound pattern in which the beginnings of tones are indicated only by very short silent intervals—an example is indicated below with explanation.

In general, our theoretical attempt aims at examining how the auditory system determines auditory streams when sound energy is distributed in various manners in time-frequency coordinates (see Nakajima, 2008, for more about music).

2. AUDITORY DEMONSTRATIONS

In order to connect AG to music, we made a few auditory demonstrations, in which a familiar melody appears illusorily. Although it may be difficult to utilize the demonstrated illusions for real music, they provide us with information on the mechanism of melody perception, which would not be clarified otherwise. After all, music is filled with illusory and ambiguous phenomena. Musicians seem to challenge the listeners' auditory capacity so that the hidden structures of music are revealed by facilitating active listening. To hear out a melody in music is not at all a simple issue.

A melody is a concatenation of notes and rests constituting a coherent structure in time. If 700-ms pure tones of 1046.5, 1174.7, and 1318.5 Hz are property connected in time, we hear a familiar melody, for example, "CDEC|CDEC" (Figure 1a). The correspondence between the physical sounds and the perceived melody is simple and clear in this case.

A 2000-ms glide component from 1046.5 to 4186.0 Hz (with a rise time of 10 ms and a fall time of 50 ms), moving at a constant speed on a logarithmic frequency scale, is perceived as a tone whose pitch ascends, but the onset of this tone can be heard as if it were an independent tone burst (Bregman, 1990). The pitch of the perceived tone burst is close to, but not exactly equal to, the pitch of a pure tone of the starting frequency (= 1046.5 Hz). The onsets of other glide components from 1174.7 to 4698.6 Hz and from 1318.5 to 5274.0 Hz can be also heard separately with pitches loosely corresponding to the starting frequencies. If these glide components are presented so that their onsets mark the beginnings of the notes in the above melody, inevitably with temporal overlaps, a melody (as in Figure 1a) with somewhat unstable pitches appears (Figure 1b).

The melody typically sounds staccato, as if played in pizzicato, indicating that the onsets are indeed separated perceptually. The same type of onset separation takes place also when descending glide components are used (Figure 1c). The separated onsets, as they are close to each other in time and frequency, are integrated perceptually, as if they were independent sounds constructing a coherent melody. The melody is an auditory stream in which each onset is followed by a silence, and each silence, except for the last one, is followed by a next onset as allowed by AG.

When a 2000-ms ascending glide component from 261.6 to 1046.5 Hz and another 2000-ms component from 1046.5 to 4186.0 Hz are presented successively (with rise and fall times of 10 ms), but with an overlap of 300 ms, we typically perceive a short tone corresponding to the overlap duration. In the present demonstration, the first component increases its level by 20 dB taking its whole duration, and the second component decreases its level by the same amount starting from the final level of the first component. This manipulation is to make the illusory tone clearer, and is not essential. Although only two 2000-ms glide components are presented physically, we hear a short tone obviously shorter than 1s (= 1000 ms), whose pitch is rather vague but hinting a tone chroma around that of a 1046.5-Hz pure tone. This is what we call the split-off phenomenon (Nakajima et al., 2000; Remijn & Nakajima, 2005). Our explanation is that the onset and the offset of the different glide components delimiting the overlap are connected to each other perceptually because they are close to each other in time and frequency. In this way, we obtain a grammatical auditory event consisting of an onset, a filling, and an offset in this order, which should be followed by a silence. We can generate such illusory tones of different pitches, and present them successively; the long glide components for different illusory tones overlap each other inevitably (Figure 1d). The familiar melody thus can be played with the illusory tones of the shorter apparent duration, followed by a short rest if a next note comes.

The above melody to be played includes only three pitches for C, D, and E corresponding to 1046.5, 1174.7, and 1318.5 Hz. We can first present the first three notes "CDE" in such a way that a pure tone of the corresponding frequency starts at the beginning of each note, and that it continues throughout the duration of the whole melody (with a rise time of 10 ms and a fall time of 50 ms). The duration of each note is set at 700 ms. All the other notes are indicated just by inserting an immediately preceding temporal gap of 50 ms which is preceded by a fall time (decay) of 50 ms (Figure 1e). Thus no tones are around 700 ms except those corresponding to the fourth and the eighth note, but eight notes of equal lengths appear comprising the familiar melody. We interpret the phenomenon that each note is indicated by the onset of a pure tone portion and a filling after that. Most notes, which are auditory events, seem to consist of an onset and a filling before connected to a next note as allowed by AG. The auditory system may have a mechanism to extract one or two representative auditory streams.

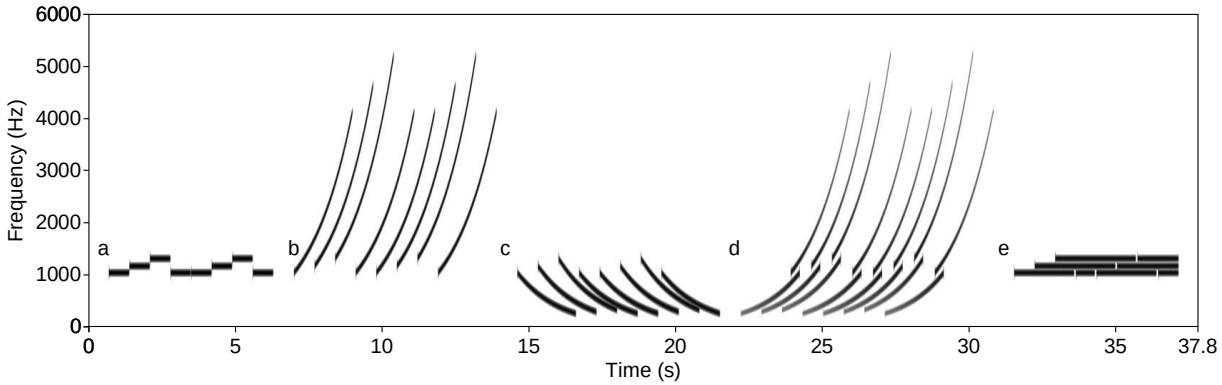


Figure 1: Spectrogram of the auditory demonstrations.

A remarkable aspect of AG is that ungrammatical sequences of auditory subevents—exactly speaking subevent cues—can be reconstructed into grammatical sequences. If a 2000-ms glide component from 261.6 to 1046.5 Hz with a rise time of 50 ms and a fall time of 10 ms is presented, the last part of this component can be extracted perceptually, with some effort, as a short tone whose pitch roughly corresponds to the ending frequency (= 1046.5 Hz). This is probably a simple example of what we call a termination tone (Nakajima, Kuroda, & Terada, 2010). The onset of the short tone is not present physically, and should be restored. The idea that an onset should be restored if AG requires has been stated and confirmed in one of our previous studies (Sasaki et al., 2008). In a recent pilot study, we were able to present several glide components so that the successive termination tones of different pitches were organized to be heard as the familiar melody (as in Figure 1a). The phenomenon of termination tones appeared more clearly in this demonstration than in previous examples. One of the reasons may have been that the filling cue before the offset of each tone was heard clearly in this example, but still this remains a remarkable phenomenon. An onset can be a complete auditory event, but an offset itself cannot be an auditory event in AG. The latter requires a preceding onset followed by a filling in order to construct a grammatical auditory event: an onset, a filling, and an offset in this order to be followed by a silence. This should make the termination tones subjectively longer than the perceptually separate onsets (as in Figures 1b and 1c), which seemed indeed the case in our pilot study.

3. IMPLICATIONS

The grammatical system developed to explain auditory organization for very simple sound patterns works also for music materials. It has become clear that music should be constructed in our mind (or brain) which seems to be constrained by AG.

4. ACKNOWLEDGMENT

This study is supported by a Grant-in-Aid (25242002 to YN in FYs 2013-2017) from the Japan Society for the Promotion of Science.

5. REFERENCES

Bregman, A. S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound, Cambridge, MA: MIT Press.

Nakajima, Y. (2008). Illusions related to auditory grammar: Ten demonstrations in musical contexts. Proceedings of the 10th International Conference on Music Perception and Cognition, Sapporo, Japan, 301-304.

Nakajima, Y., Kuroda, T., & Terada, T. (2010). Auditory events that seem to be formed illusorily from offset cues. Proceedings of the 26th Annual Meeting of the International Society for Psychophysics, Padua, Italy, 245-250.

Nakajima, Y., Sasaki, T., Kanafuka, K., Miyamoto, A., Remijn, G., & ten Hoopen, G. (2000). Illusory recouplings of onsets and terminations of glide tone components. Perception & Psyhchophysics, 62, 1413-1425.

Nakajima, Y., Sasaki, T., Ueda, K., & Remijn, G. B. (2014), Auditory Grammar, Tokyo: Corona Publishing (in Japanese).

Remijn, G. B. & Nakajima, Y. (2005). The perceptual integration of auditory stimulus edges: An illusory short tone in stimulus patterns consisting of two partly overlapping glides. Journal of Experimental Psychology: Human Perception and Performance, 31, 183-192.

Sasaki, T., Nakajima, Y., ten Hoopen, G., van Buuringen, E., Massier, B., Kojo, T., Kuroda, T., & Ueda, K. (2010). Time-stretching: Illusory lengthening of filled auditory durations. Attention, Perception & Psychophysics, 72, 1404-1421

OBSERVERS CAN RECOGNISE REAL VERSUS FAKE FREE-IMPROVISATION DUOS THROUGH NONVERBAL BACKCHANNEL CUES

Nikki Moran[1], Lauren Hadley[1,2], Peter Keller[3,4]
[1]Institute for Music in Human and Social Development, Reid School of Music, University of Edinburgh, United Kingdom, [2]Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, United Kingdom, [3]The Marcs Institute, Music Cognition and Action Group, University of Western Sydney, Australia, [4]Max Planck Institute for Human Cognitive and Brain Sciences, Research Group: Music Cognition and Action, Leipzig, Germany
Time: 12:00

Joint improvisations require performers to integrate their musical contributions so that audience members perceive a single event, rather than two individual performances. Non-musical interaction studies show that observers can recognize true affiliation between conversing dyads under sparse conditions (e.g. silent video of interactors, and where motion-captured representations are devoid of accompanying facial expression). Observers thus appear to make judgments based on the contingency of ostensive listeners' nonverbal 'backchannel' behavior. This study aimed to examine: 1. Whether observers can reliably identify the authenticity of genuine and fake (mismatched) duos in a musical interaction scenario; and 2. Whether musical background and individual rhythm perception skills are associated with task success. 24 improviser musicians were recruited for twelve (duo) recording sessions. The full study involved both standard jazz and free-improviser duos. In this paper we report only findings from the 6 free-improviser duos. The musicians performed short improvisations, following instructions to include brief, unaccompanied solo sections. A Vicon Nexus 1.6.1 system was used to capture and model 3D motion (10 cameras at 200Hz; 18 light-reflective markers per individual). Stereo audio was recorded separately. Eight 10-second point-light representations (.mov files) were generated from the synchronized kinematic and audio recordings of solo sections, in which one musician is completely silent. Eight further representations were created by splicing 'soloists' with 'listeners' from different duos. 60 participants with various musical backgrounds were asked to rate as either real or fake the 16 randomized point-light movie trials in four repeated blocks. They also completed a rhythm ability test. Participants' sensitivity to the task was measured with signal detection indices. Variance was analyzed across four musical experience groups. Post-hoc regression examined the effect of individuals' rhythm skills.
We found that participants were sensitive to the real/fake distinction ($t(59)=5.05$, $p<.001$), showing no significant difference across the four musical experience groups. Individual differences in rhythm skill offer no further explanation of variance. Participants overall performed the task reliably. Musical background and rhythmic ability appear to offer no sensitivity advantage, giving no further account of variance. These results suggest that the perception of backchanneling in free-improvisation is a general ability, which may be related to some other aspect of social competence.

[10C] Expectation, Education and Evaluation in Performance

MONTREAL 11:00-12:30 Friday 8 Aug 2014

CHERISHING THE ERROR: CLASHING SCHEMAS AND COGNITIVE BIAS IN PIANO PERFORMANCE

Kristin Kueter, Blair Kaneshiro, Jonathan Berger
Stanford University, USA
Time: 11:00

ABSTRACT

Common errors in initial piano studies sometimes appear to stem from schema miscues. That is, rather than arising from any obvious physiological or motoric challenge, the error suggests a subconscious interpretation of an overt or subtle ambiguity. These miscues typically implicate a cognitive bias, often an anchoring heuristic. We polled numerous professional piano teachers, gathering examples of commonly recurring errors amongst their students, as well as descriptions of how they approached these errors in their teaching. We present examples of their responses along with our analysis of these occurrences. We propose a seemingly rarely taken approach of addressing these situations not as 'errors' but as ambiguities to be appreciated by the student, hence 'cherishing the error.' Finally, we consider evidence of cognitive biases in interpretive decisions of professional performers.

1. INTRODUCTION

Schemas optimize mental processing and are essential to expectation formulation. Occasionally, however, schema selection might impede appropriate, and even vital, interpretations. They may "prevent us from seeing what would otherwise be obvious" (Sweller, 2011). Such 'miscues' are prone to cognitive biases that may result in particular judgments about an ambiguous context through an auditory analog of the focusing effect, or due to an anchoring heuristic (Vass, 2014).

In this paper, we consider instances of apparent schema miscues in piano performance and consider the ensuing cognitive biases in terms of resulting behavioral heuristics. We consider instances in which schema miscues might affect both learning and recall, particularly by children. We propose that appreciating the underlying ambiguity of the miscue (i.e., 'cherishing the error') might afford a valuable teaching opportunity, not only about music, but also about cognitive bias in decision making.

2. CASE STUDIES

Consider the opening measures of the Musette in D from *Clavierbüchlein der Anna Magdalena Bach* (BWV Anh 126), a canonical pedagogical piece in the piano repertoire. Results from our survey of piano teachers consistently emphasized the tendency for students to accentuate the octave As in measures 3 and 4 (as shown in Figure 1), effectively turning the two sixteenth notes in measure 3 into an upbeat. This engenders a metric shift by a single eighth note. The student typically continues the shifted meter, accenting the second eighth note of measure 4.

The durational accent on the A (second eighth note in measure 3), the ascending line to that A, the descending conjunct tetrachord that follows, and the registral return to the A on the second eighth note of measure 4 all articulate multiple schemas that collaborate to create this abrupt metric shift – particularly remarkable given the fact that the meter has just been so strongly established in the preceding measures.

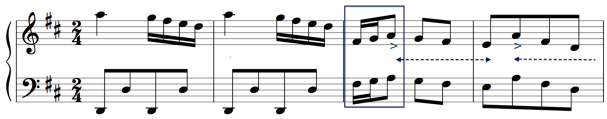


Figure 1. Measures 1-4 of the Musette in D from *Clavierbüchlein der Anna Magdalena Bach* (BWV Anh 126). The two sixteenth notes in measure 3 are often played as upbeats to the descending conjunct tetrachord that follows.

Piano teachers we interviewed consistently report this 'error', and describe methods of 'correction' that primarily comprise added visual cues (as in Figure 2). This is one example of numerous amassed anecdotal responses relating instances in particular works in the standard piano pedagogy repertoire in which student performance errors seem to result from a musical schema contradictory to the notated score.

In our study, we isolated instances in which such 'miscues' were reported as commonly occurring amongst a broad population of students by a wide variety of music teachers. While typically the teacher will simply attempt to correct the 'error', we suggest that such schema miscues often pose creative teaching opportunities to discuss the ambiguity and elucidate the cognitive bias that fosters an 'interpretation', rather than merely flagging the performance as 'erroneous'.

Experientially acquired schemas are key in context creation. Competing schemas in music can include visual assumptions. For example, the spatial schema (Tversky, 2001) of the space between delimiting bar-lines represents an accentually initiated metrical unit. Students are taught early on to 'count beats of the measure.' In Figure 1, these visually premised, metrical expectations do not correspond with the auditory schema driving the student's performance.

Instances such as these are remarkable in that the schematic expectation is often so firmly felt that the spatial schemas in the score—including bar lines, beaming and slurring—as well as teacher's annotations—such as arrows, circles, exclamation marks, and other devices—are not enough to override students' schematic musical expectations.

2.1. Miscues and Cognitive Biases

We culled examples of piano performances by children with notable schema miscues. These were collected from a sampling of experienced piano teachers who, collectively, have taught more than 2,000 students over the past 50 years. In interviews with these teachers, we tried to follow the process of their attempted 'correction' of the 'error'. We noted cases in which behavioral anchoring heuristics are at play in the student's interpretation and execution of the excerpt, and discussed the teaching method of addressing these moments.

In all of the examples collected, the students' schematic expectations conflict with the notated music. In some cases the inherent ambiguity between what is notated and the schema that is seemingly invoked seems to be intentional on the part of the composer, while in other cases, the ambiguity seems to be subtler, if not entirely unintentional. The child often finds it difficult to rectify the 'erroneous' interpretation as the schema miscue creates a behavioral bias, most often an anchoring heuristic (Tversky & Kahneman, 1974).

Figure 2 presents a marked score from one of the participating piano teachers. The work, Olympics Parade by James Bastien, is a commonly used pedagogical piece for children. The 'error' apparently results from the student internally hearing the opening measure as a dominant preparation to a major chord in Measure 2 and, understandably, plays without regard to the notated key signature.

The level of frustration on the part of one teacher in trying to correct the child's persistent erroneous F-natural is evident in the penciled corrections in Figure 2. As the teacher describes, the 'error' often persists even after the key signature is circled and the accidentals marked in multiple colors on the score.

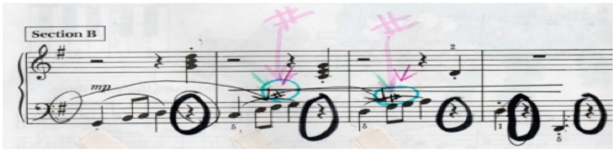


Figure 2. A teacher's notated corrections in a reportedly common performance error in Olympics Parade, a pedagogical piano work by James Bastien. The student has interpreted the tonic G-Major chord in measure 1 as the dominant of C Major, resulting in F-natural (rather than F-sharp) being played in measure 2.

A similar error, again a mishearing of the incipient tonic as a dominant, is evident in the commonly recurring error near the opening of the Minuet in G from *Clavierbüchlein der Anna Magdalena Bach* (BWV Anh 114**)**, shown in Figure 3.

As in the previous example, many students seem to hear the opening as a dominant upbeat heading toward the perceived structural accent of measure three. The (mis-)interpretation of the opening measure as a dominant rather than a tonic appears firmly anchored, overriding the plainly visual cue of the key signature. Again, disregarding the key signature, students anchor the piece in the key of C Major rather than G Major.

While erroneous, this interpretation is not irrational. The prolongation of the incipient harmony through the second measure reinforces the interpretation of a dominant preparation. Simply referring the student to the key signature constitutes a 'correction', but does not recognize the musical rationale behind the error. Thus, in our study we choose to designate instances such as these as 'miscues' rather than 'mistakes'. The miscue has validity, although it is based upon a false assumption resulting from anchoring to the strength of what they are hearing and seeing in that very first measure.27

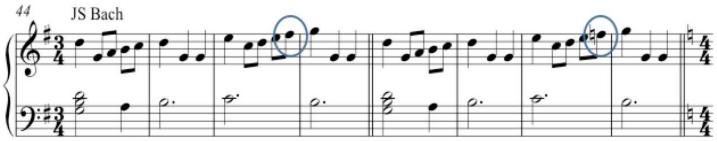


Figure 3. Measures 1-4 of *Minuet in G Major* (JS Bach/C. Petzold, *Minuet, Clavierbüchlein der Anna Magdalena Bach,* BWV Anh 114). The piece as notated (left) and as performed (right).

Heuristic behaviors come into play in economic situations either because people have been given an errant starting point, or because they lack the time needed to reach a fully informed decision. This is similar to how a musical performer might improvise his way through sheet music because it sounds right even though it might not be true to the composition. In music this is called 'faking it'; in economics it is a cognitive bias that is based on a gut instinct. But in both cases, those choices are not random. In music, performers are given general performance guidelines through knowledge of beat and key signatures. In economics, something is also anchoring your decision—either someone is giving you a starting number or you are surmising it from suggestion.

One such guiding behavior is the Anchoring Heuristic, first written about by Amos Tversky and Daniel Kahneman in their seminal 1974 work, *Judgment under Uncertainty: Heuristics and Biases*, and this principle comes into play in both musical miscues and economic cognitive biases. They state:

> *In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient. That is, different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring.*

These anchoring situations can occur when "a starting point is given to a subject," or "when the subject bases his estimate on the result of an incomplete computation" (Tversky & Kahneman, 1974, p.1128). An example of the former: A used-car salesman spots a buyer and tries to sell her a car for $1,000, knowing full well that its worth is only $500. The dealer has anchored the buyer with that $1,000 price, forcing her to negotiate down from that point, this despite the fact that the dealer could have started at any number. Now, suppose the buyer negotiates the dealer down to $750 for the car. She leaves the lot feeling content in her conviction that she has purchased at a good price, based upon its reduction from the starting price that was given.

In both musical performance errors and economic cognitive biases, these heuristics are not limited to the young or inexperienced. As Tversky and Kahneman (1974, p.1130) point out, "The reliance on heuristics and the prevalence of biases are not restricted to laymen. Experienced researchers are also prone to the same biases when they think intuitively."

In a casual study, we asked the very piano teachers who corrected the 'error' in Figure 1 to vocalize the melody. In many instances, the teachers fell trap to the same 'error' they had been correcting in their students for decades: While they consistently corrected the error in pedagogical situations, they themselves performed it when asked to sing or hum the passage. Furthermore, some insisted they had reproduced the section correctly, not realizing their error until we pointed it out to them.

Beyond pedagogy, a survey of recorded performances—both by professional and amateur musicians—reveals the inherent ambiguity, with many, if not most (including Eugene Ormandy's performance of an orchestral arrangement with the Philadelphia Orchestra28), performing what most piano teachers earmark as 'erroneous'.

3. DISCUSSION

Labeling a given interpretation of a musical ambiguity as 'right' or 'wrong', with preference given to visual versus musical schema, may have pedagogical rationale in terms of imparting literality. However, the remarkable fact is that in instances such as that of the *Musette*, the intuitive interpretation of the child—based upon schematic expectations (and, due to the anchoring effect, difficult to be dissuaded of) is retained in both the intuitions of the teacher (when 'caught off guard'), as well as the conscious performance decisions of mature artists. There is potential for a valuable, profound teaching moment in recognizing the ambiguity.

Beyond piano pedagogy, placing heavy emphasis on the first piece of information we encounter can affect our subsequent decisions– for better or worse—and heuristics cross over into many domains of our lives. Whether learning the piano or shopping for a used car, the mechanisms behind the underlying heuristics are essentially the same.

The argument in favor of one tonal or metric interpretation over another constitutes a rich but often avoided aspect of piano pedagogy and music education. As we proceed with behavioral studies of schema miscues and cognitive biases, we seek to further understand how age affects interpretation. Is it possible (and desirable) to develop an instructional curriculum that develops an appreciation of the process of making interpretive choices from early on, rather than presuming that maturity, skill, and sophistication are prerequisites for such choices?

27 The authors would like to add that the children correctly play F-sharps in other places within the composition, so this is not a result of physiological constraints or motor coordination in an attempt to play the black key. And again, this is an error that is repeated over decades by numerous students with different teachers.

28https://itunes.apple.com/us/album/baroque-masterpieces-pachelbel/id173975441, Track 5

Some methodologies (the Suzuki method, for example) inherently avoid recognizing ambiguity; other approaches favor notation over intuition. Most methods advocate mimicry. The potential pedagogical rewards of 'cherishing the error' not only fosters deeper appreciation of music but also could teach children to recognize and navigate the judgmental complexities imposed by cognitive biases.

4. SUMMARY

By examining musical performance errors from an economic heuristic perspective, we show that these might not be errors at all but creative, intuitive judgment calls or 'miscues' based upon firm principles. Therefore, it may be more useful for educators to discuss with their students the anchoring behind a miscue rather than simply labeling and squelching the 'error'. Understanding the underlying mechanism behind these miscues and reconciling the ambiguity can foster creativity and constitute a more insightful and lasting pedagogical experience.

5. REFERENCES

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory. *New York: Springer.*
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), *1124-1131.*
Tversky, B. (2001). Spatial schemas in depictions. In Spatial schemas and abstract thought *(pp. 79-111).*
Vass, Z. (2012). A psychological interpretation of drawings and paintings. The SSCA method: A systems analysis approach. *Alexandra publishing.*

EVOLUTION OF TIMING THROUGH INSTRUMENTAL PRACTICE BY VIOLINISTS AND CLARINETTISTS.

Dirk Moelants[1], Luc Nijs[1], Marc Leman[1]
[1]Ipem, Dept. of Musicology, Ghent University, Belgium
Time: 11:30

Timing is one of the most studied topics in music perception research. Tapping tasks are commonly used to study perception and motor skills, and timing in music performance is considered as a prime factor in expressivity. Despite this, the evolution of timing skills from beginning musicians to professional players has rarely been studied. One aspect of particular interest is the effect of the musical instrument played. In this study the evolution of timing by violin and clarinet students are compared. These instruments use completely different techniques and as such require different skills from the performers. Do these difference also lead to a different sense of timing? And do these differences persist with increasing skill level or do they gradually disappear? Four simple rhythms were presented using a synchronization-continuation paradigm: a simple regular pulse at 120 bpm, a dotted pattern (3:1), a long-short-short pattern (2:1:1) and an irregular '8/8'-pattern (3:3:2). 100 musicians participated, half of them violists, half clarinettists, divided in 5 skill levels from beginners to professionals. Their performances were recorded and the timing was analyzed in terms of synchronization, tempo and micro-timing. A clear evolution of timing is visible for both instrument groups. The timing becomes more regular and synchronization more precise. Already with beginning players we see differences between the two instruments, clearly related to technical difficulties. Thus, clarinettists tend to slow down in order to regulate their breathing, while violinists tend to play the dotted patterns less 'sharp' than clarinettists. Interestingly, some of these elements of micro-timing persist and are found even with advanced performers, only disappearing at the highest professional level. This study gives a view on how timing evolves during the formation of musicians. It also shows that the choice of the instrument has an effect on the timing of professional musicians.

A COMPUTATIONAL MODEL OF IMMANENT ACCENT SALIENCE IN TONAL MUSIC: PERCEPTUAL EVALUATION

Florian Eckl[1], Erica Bisesi[1], Anders Friberg[2], Richard Parncutt[1]
[1]University of Graz, Austria, [2]Kth, Stockholm, Sweden
Time: 12:00

ABSTRACT

Bisesi, Friberg and Parncutt (2014) developed a semi-algorithmic approach to music performance rendering estimating the location and perceptual salience of immanent accents (grouping, metrical, melodic, harmonic) in the musical score. In this study, we test this algorithm by comparing predictions with listeners' ratings. 12 expert musicians hear 8 contrasting recordings of each of two Mozart Duets (Sonata K 448, 3rd movement and Sonata K 521, 2nd movement) and select salient events by clicking on the screen as the music progresses. They then hear each performance again and rate the importance (salience) of metrical, melodic and harmonic accents predicted by the model. By systematically comparing the treatment of immanent accents by different performers, the model is being gradually updated to optimize the agreement between predictions and data.

1. BACKGROUND

Parncutt, Bisesi and Friberg (2013) developed a semi-algorithmic approach to music performance rendering estimating the perceptual salience of immanent accents (grouping, metrical, melodic, harmonic) in the musical score.

The approach of that study is based on the analysis-by-synthesis approach of Sundberg (1988) and Friberg (1991), and their rule-based performance rendering system *Director Musices* (DM). Inspired by Sundberg and Friberg, Parncutt (2003) developed a new theoretical foundation. In a broad definition, an *accent* is any musical event that seems important or attracts the attention of a listener. An *immanent* accent is an accent that is determined by the musical structure as suggested by the musical score; a *performed* accent is added to the music by the performer by manipulating dynamics, timing, articulation or timbre. In both cases, the perceptual *salience* of an accent is its perceptual or subjective importance, or the degree to which it attracts a listener's attention.

Many of the performance rules of Sundberg and Friberg can be reinterpreted in terms of this general concept of accent. The accent concept unifies the theory under a new general umbrella and establishes a stronger connection between performance rendering and the academic discipline of music theory and analysis.

Grouping (or phrasing) and metrical accents emerge from the hierarchical structure of phrasing and meter, melodic accents follow melodic leaps and are strongest at contour peaks and (to a lesser extent) valleys, and harmonic accents depend on vertical and horizontal (simultaneous and successive) aspects of dissonance including surprising chord/key changes.

Bisesi, MacRitchie and Parncutt (2012; 2013) examined the perception of phrasing boundaries, climaxes and accents in different performances for pieces by Chopin (solo) and Mozart (duet), comparing these with model predictions for immanent accent positions.

2. AIMS

We test an updated version of our algorithm of immanent accent salience by comparing predictions with listeners' ratings.

3. METHOD

3.1. Participants

12 expert musicians are recruited from the University of Graz and from the KTH, Stockholm.

3.2. Materials

Excerpts lasting about 50 sec are taken from 8 contrasting commercial recordings of two Mozart pieces for piano duo (Sonata K 448, 3rd movement) and four hands (Sonata K 521, 2nd movement). The 8 performer duos are Martha Argerich and Evgeny Kissin (recorded the pieces in 2001 and 2004 respectively), Güher and Süher Pekinel (1989 and 1991), Yaara Tal and Andreas Groethuysen (both pieces in 2005), Ingrid Haebler and Ludwig Hoffman (both pieces in 1978), Ennio Pastorino and An Li Pang (1989 and 1988 respectively), Benjamin Britten and Sviatoslav Richter (1967 and 1966), Martha Argerich and Alexandre Rabinovich (both in the period 1992-1993), and Cristoph Eschenbach and Justus Frantz (both pieces recorded in 1972).

3.3. Procedure

Participants are asked to perform two tasks.

The first tasks focuses on accents' locations. Participants listen to the performances in random order and indicate for each performance when an important event is perceived by clicking on the screen. A Java interface specifically developed for the purpose of this study associates to each mark the corresponding instant of occurrence in the performance.

All the selected events are interpreted as accents of a given kind (i.e. metric, melodic, or harmonic).

In a second stage, we check model predictions concerning accent salience. Accent analysis has been performed for each piece identifying immanent accents in metric, melodic, and harmonic categories. Metric and melodic accents are estimated in Director Musices (Bisesi and Parncutt, 2011; Parncutt, Bisesi and Friberg 2013); harmonic dissonance accents are estimated by counting the rate of occurrence of each pitch class set in each piece by mean of the Humdum toolkit (Huron, 2002; Sapp, 2011), and comparing these rates with the familiarity of pitch class sets in a database including 18th and 19th century vocal polyphony; surprise accents are computed by combining pitch class sets theory and Parncutt (1988) psychoacoustical model of musical chords' roots. In this task, theoretically predicted accent locations are presented to participants as markers in the timeline of each performance. As the music progresses, participants are asked to indicate the perceived salience at each marker position, by clicking on a vertical grid corresponding to a rating scale from 1 to 3.

For each of the two stages, participants have 2 trials to become familiar with the procedure.

The total duration of the experiment is about 30 min.

4. PRELIMINARY RESULTS AND DISCUSSION

A pilot study on two Chopin Preludes (Bisesi, MacRitchie and Parncutt, 2012) suggested that, on average, phrasing, metrical, melodic and harmonic accents are about equally salient; and subjective phrase climaxes correspond to predicted accents with a small time lag. In a second pilot study involving the two Mozart Duets (Bisesi, MacRitchie and Parncutt, 2013), we found passages of music where accents were rated differently in different performances, reflecting idiosyncrasies in each piano duo's interpretation.

There is a general qualitative agreement between predictions and data. In the current stage of the analysis, the model is being gradually updated to optimize the agreement between predictions and data. In a further stage, we will also consider differences and similarities among performances, and attempt to cluster them according to performers' styles.

5. ACKNOWLEDGEMENTS

This research is supported by the FWF Stand-Alone Project P 24336-G21 "Expression, emotion and imagery in music performance". We kindly thank Dieter Kleinrath (KUG, Graz) for his support in the implementation of the computer interface used in this study.

6. REFERENCES

Bisesi, E., & Parncutt, R. (2011). An accent-based approach to automatic rendering of piano performance: Preliminary auditory evaluation. Archives of Acoustics, 36 (2), *283-296.*

Bisesi, E., MacRitchie, J., & Parncutt, R. (2012). Recorded interpretations of Chopin Preludes: Performer's choice of score events for emphasis and emotional communication. In E. Cambouropoulos et al. (Eds.), Proceedings of the ICMPC-ESCOM 2012, *Thessaloniki, Greece, 2012 (pp. 106-107).*

Bisesi, E., MacRitchie, J., & Parncutt, R. (2013). Structural communication in piano duos: Musical compatibility and individual differences in interpretation. In A. Williamon and W. Goebl (Eds.), Proceedings of ISPS 2012, *Vienna, Austria, 2013.*

Friberg, A. (1991). Generative rules for music performance. Computer Music Journal, 15 (2), *56-71.*

Huron, D. (2002). Music information processing using the Humdrum toolkit: Concepts, examples, and lessons. Computer Music Journal, 26, 2, *11-26.*

Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord, Music Perception, 6, *65-94.*

Parncutt, R. (2003). Accents and expression in piano performance. In K. W. Niemöller (Ed.), Perspektiven und Methoden einer Systemischen Musikwissenschaft *(pp. 163-185). Frankfurt/Main, Germany: Peter Lang.*

Parncutt, R., Bisesi, E., & Friberg, A. (2013). A preliminary computational model of immanent accent salience in tonal music. In Proceedings of SMAC-SMC 2013, *Stockholm, Sweden, 2013.*

Sapp, C. S. (2011). Computational methods for the analysis of musical structure. PhD Dissertation.

Sundberg, J. (1988). Computer synthesis of music performance. In J. A. Sloboda (Ed.), Generative processes in music. *Oxford: Clarendon.*

[10D] Singing 2

LIÈGE 11:00-12:30 Friday 8 Aug 2014

AUTOMATIC IDENTIFICATION OF EMOTION IN CHINESE OPERA SINGING

Dawn Black[1], Ma Li[2], Mi Tian[1]
[1]Queen Mary University of London, United Kingdom, [2]Central Academy of Drama, China
Time: 11:00

ABSTRACT

An emotional cue is understood to be a property of sound that evokes an emotional response in a listener. This paper considers a novel and un-researched problem: that of automatic identification of emotional cues in singing. An important contributor to the overall perception of a piece of music, singing and the use of voice as an instrument has been largely ignored by the Music Information Retrieval community. In this work we view the vocal signal from the perspective of a speech signal and follow the approach adopted by this community. We extract low-level features commonly used to detect emotion in speech (and sometimes in music) and use a Support Vector Machine to classify our songs according to emotional content. A good classification rate identifies the feature as an emotional cue. We show that a number of the extracted feature sets perform well when differentiating emotion in singing. The MFCCs, formants (Burg) and combined spectral descriptors are identified as the best performers.

1. INTRODUCTION AND RELATION TO PRIOR WORK

The voice is the oldest musical instrument and one that we all possess. Singing therefore plays an important role in nearly all music, to the extent that much popular music is known by the singer's name (artist's name) only. Despite its prominence, it has been little considered by researchers in Music Information Retrieval (MIR) or emotion recognition in audio. When viewed as the crossover between music and speech, a rich research area with much potential becomes apparent. While not being the same as either speech or music, we would expect unaccompanied singing to contain some attributes from both areas, and some unique to itself.

The communication of human emotion is the topic of much research and discussion. How humans encode (expression) and decode (impression) emotion has been much analysed (Scherer 2003). To date, research into emotional encoding in audio signals has concentrated primarily on speech and music in general, but little research has specifically addressed the unique properties of the singing voice. This new viewpoint is one of the major points of novelty for this paper.

Singing can be considered the intersection of music and speech and therefore it is sensible to combine existing approaches in these two areas when searching for emotional cues in singing. However, we expect significant differences. For example, pitch is one of the primary features used for emotion recognition in speech but is not considered useful for music (Lu, Liu & Zhang, 2006). Is it an emotional cue in singing? The remainder of this section positions our work within these two fields, and refers to existing work in the area of singing.

Emotion recognition in speech is a popular research area with many notable publications. In (Juslin & Laukka, 2003) Juslin et al. present an excellent table that shows how longstanding and widely researched this field is. The fundamental principles of emotion recognition are well established and simple - extract some features that are considered to contain emotional cues from the speech signal and then use a classifier to group samples of similar emotional content. A good review of features and classifiers, and their effectiveness can be found in (Ververidis & Kotropoulos, 2006). Features extracted are many and the community at large has yet to converge on a standard feature set.

Emotion (or mood) recognition in music takes a similar approach (feature extraction followed by classification) but the extracted features are different due to the differing nature of the audio. Low-level features common to emotion recognition in speech can be employed, such as MFCCs, zero-crossing rate and pitch, but many high-level features, such as key and beat tracking, are particular to music (Pohle, Pampalk & Widmer, 2005).

Davis (Davis, 1998) and Tsai et al. (Tsai, Wang, Wang, Shau, Hsiao & Auhagen, 2009) discuss the topic of emotion encoding in singing at length but feature extraction is not performed. In other papers (Siegwart & Scherer, 1995; Rapoport,1996; Sundberg, Iwarsson & Hagegaård, 1995) results mostly rely on listening tests or visual examination of spectrograms. This time-consuming approach constrains the song database and results, dependent as they are on human input only, can be unconvincing.

The extension of automatic feature extraction and classification to singing is one of the major contributions of this paper. It enables singing voice analysis to progress from the manual inspection of a small selection of singers and pieces to automatic analysis of very large data-sets. Work in this area is sparse. In (Goto, Saitou, Nakano & Fujihara, 2010) Goto et al. presented a review of different singing research being conducted by their group, including MIR based on singing (Fujihara & Goto, 2007) where a system for MIR based on vocal timbre is presented.

In this paper we apply the principals of automatic emotion recognition in speech and music to singing. We extract a set of features common to speech and some low-level features common to music from an unaccompanied singing database that contains samples identified as having either positive or negative emotion. A support vector machine is employed to classify the songs according to emotional content. Results show that this approach is successful at extracting and identifying emotion in singing.

2. SINGING DATABASE

When constructing a database one must consider the approaches adopted by both the speech and music communities. The speech research community will record a number of speakers repeating the same utterance but simulating different emotions. The aim being to rule out all potential sources of variation other than the emotion portrayed. This approach is not suitable for music or singing. Songs and musical pieces are learned and practiced with a specific emotion intended. It then becomes very hard to perform a familiar piece but portray an alternative emotion. Although not inconceivable, a performer could not be expected to portray the alternative emotion with the same level of skill as the learned emotion. We have therefore taken the approach adopted by the music community where different emotions are portrayed by different pieces.

The difference between spontaneous, naturally occurring emotion and acted emotion must also be mentioned. When constructing an emotional speech database, speakers simulate or act an emotion. The differences between simulated and spontaneous emotion in speech are a topic of much discussion. However, when considering singing, as with music, all emotion is simulated.

When analyzing any audio signal with musical properties the problem of genre arises. If the database consists of samples from many, varied, musical genres we risk confusing emotion and genre recognition. However, if all the samples come from the same genre we risk limiting ourselves to musical expression within that genre (eg. rock tends to be predominantly aggressive). Finally emotional content is subjective. The aim is therefore to select a genre that is wide enough to encompass a broad range of emotions, is well established and has well-defined, recognised modes of expression and is popular enough to have a record of being interpreted similarly by a large audience from varying backgrounds.

Chinese Opera fulfils all these criteria. It has existed for hundreds of years and so the emotional content of songs and mode of expression are well established. Additionally the songs all form part of a play or story so the emotion expressed fits a well-defined situation. This results in a clear consensus about the principal emotion the singer intends to convey. Much of the research already in the area of singing analysis uses Opera (Rapoport, 1996; Sundberg, Iwarsson & Hagegaård,1995; Siegwart & Scherer, 1995) for similar reasons. Some analysis of Chinese Opera has been published (Zhang, Zhou & Wang, 2008; Qu & Liu, 2006; Dong, Sundberg & Kong, 2014; Dong, Kong & Sundberg, 2013; Sundberg, Gu, Huang & Huang, 2012) but none in the area of emotion recognition.

A database of 15 unaccompanied singers, all of whom had undertaken significant formal training, was recorded. Singers were asked to provide songs expressing both positive and negative emotion. The performer was asked to describe the emotion of the song and this was considered ground truth for our experiments. Songs were then segmented into shorter phrases between 3 and 6 minutes long. This produced a database of 516 audio samples. All recordings were unaccompanied singing sampled at 44.1 kHz or higher. Samples were normalised in amplitude and reduced to mono. This database is available for download under a creative commons license at http://c4dm.eecs.qmul.ac.uk/rdr/ all usage should cite this paper.

3. FEATURE EXTRACTION

This section presents details of the features used. Features were extracted using either Matlab or PRAAT (Boersma & Weenink, 2012). Although PRAAT is designed for speech analysis, it is a fairly simple matter to adapt the PRAAT settings for measurement of singing by increasing the maximum expected frequency from one suitable for speech to one suitable for singing. For these experiments the upper bound on frequency was set to be 1050 Hz which is just higher than C_6 or the soprano *C*. All features were calculated for a frame size of 10ms.

3.1. Silent, voiced and unvoiced frames

Speech and unaccompanied singing typically contain periods of silence (less common in music). Feature extraction from silent frames will be unreliable and so silent frames were discarded. Speech and unaccompanied singing also contain unvoiced utterances which exhibit no periodicity. The extraction of periodicity based features is not appropriate for unvoiced frames and hence these features were extracted for voiced frames only.

3.2. Pitch and related features

In order to detect the pitch we use the YIN algorithm (Cheveigne & Kawahara, 2002) and 10ms voiced frames. The mean, median, variance, maximum and minimum for the pitch vector and first and second pitch vector derivatives were calculated to give a 15-dimensional pitch vector. The use of the first derivative is fairly standard in speech communities. The use of the second derivative is less common but is predominantly applied to tonal language analysis (Yeh, Pao, Lin, Tsai & Chen, 2010).

3.3. Audio power

A 6-dimensional power vector was calculated for each sample. The mean power, variance, minimum, maximum, mean voiced frame power and mean unvoiced frame power.

3.4. Harmonic to Noise Ratio (HNR)

The HNR is a measure of the ratio between the harmonic and non-harmonic components in the power spectrum. It is used to indicate the level of vocal hoarseness and hence is a good indicator of emotion (Tsai, Wang, Wang, Shau, Hsiao & Auhagen, 2009). Defined in (Boersma, 1993) as

$$HNR\ (dB) = 10log_{10}\frac{\acute{r}_x(\tau_{max})}{1-\acute{r}_x(\tau_{max})}$$

where $r'_x(\tau_{max})$ is the local maximum of the normalized correlation. A seven-dimensional HNR vector consisting of the HNR mean, variance, minimum, maximum, HNR ratio of unvoiced to voiced frames, and the average duration of voiced and unvoiced periods was created for each singing sample.

3.5. Mel-Frequency Cepstral Coefficients (MFCCs)

MFCCs are the most popular spectral representation of speech and they have been shown to achieve good speech emotion classification results (Ayadi, Moataz & Kamel, 2011). The method for calculation of the MFCCs is well known and so will not be detailed here. We extract a 13th order coefficient vector representing the spectral power in each critical band of the Mel-filter and a 13th order vector giving the spectral bandwidth in each filter. For each coefficient and bandwidth we calculated the mean, variance, minimum and maximum; also the derivative of each coefficient and bandwidth, and its mean, variance, minimum and maximum. This gave a 208-dimensional feature vector.

3.6. Jitter and Shimmer

When considering voice, jitter is defined as the small deviations in glottal cycle length. There exists research to support the use of jitter as a measurement of emotional content in speech (Wang, Li & Fang, 2006) and therefore we might expect singers to utilize jitter (or shimmer) to convey emotion when singing. However, both jitter and shimmer are assumed to be involuntary perturbations brought on by stress. It is uncertain

whether singers will deliberately employ such perturbations to convey emotion. A good classification result here would confirm their deliberate use.

The general equation for jitter is

$$jitter = \frac{\frac{1}{N-1}\sum_{n=1}^{N-1}(|T_n - T_{n+1}|)}{\frac{1}{N}\sum_{n=1}^{N-1}T_n}$$

where *n* is the frame number, *N* is the number of frames and *T* is the fundamental period. PRAAT provides five different jitter measurements.

Shimmer is a measure of the frequent deviations in amplitude in a voice. It has been shown to provide good discrimination of emotion in both humans and animals (Li, Tao, Johnson, Soltis, Savage, Leong & Newman, 2007).

The general equation for shimmer is

$$shimmer = \frac{\frac{1}{N-1}\sum_{n=1}^{N-1}(|A_n - A_{n+1}|)}{\frac{1}{N}\sum_{n=1}^{N-1}A_n}$$

where *A* is the waveform amplitude.

3.7. Formants

A formant is a concentration of energy at one or more frequencies in voice and reflects the natural resonances of the vocal tract. The position and identification of formants for vowel identification in speech is well researched but has not been investigated for singing. Their variation due to emotional state in speech, although well documented, is not certain. Sources such as (Petrushin, 2000) suggest that the formants may contain some information about emotional content but others, such as (France, Shiavi, Silverman, Silverman & Wilkes, 2000) point out that results in different works often contradict. This is primarily attributed to the difficulty in correctly identifying the formants (Ververidis & Kotropoulos, 2006) as often the first formant is confused with the fundamental frequency or the second formant.

PRAAT provides four different formant measurements; the Burg, the robust, the keep all and the split-Levinson procedures. The Burg procedure uses the Burg method for calculation of the LPC coefficients as opposed to the autocorrelation method used for the robust procedure. Results are not presented for the split-Levinson procedure as it does not provide formant bandwidth.

3.8. Low-level MPEG-7 Descriptors

Spectral descriptors are used as an indication of timbre in mood detection in music (Kim, 2010; Laurier, 2007) and speech (Lampropoulos & Tsihrintzis, 2012). We have taken a selection of low level spectral descriptors described in the MPEG-7 standard (ISO-IEC 15938).

Low level spectral descriptors provide a compact description of the signal spectral content. The use of logarithmic frequency scales is an approximation of the response of the human ear. The Audio Spectral Envelope (ASE) is an indicator as to whether the spectral content of a signal is dominated by high or low frequencies. The Audio Spectral Centroid (ASC) descriptor could be considered as an approximation of perceptual sharpness of the signal. The Audio Spectral Flatness (ASF) describes the flatness properties of the spectrum of an audio signal for each of a number of frequency bands. The Audio Spectrum Spread (ASS) descriptor indicates whether the signal content is concentrated around its centroid or spread out over a wider range of the spectrum. This gives a measure which allows the distinction of noise-like sounds from tonal sounds.

The ASE was calculated across 34 frequency bands and the mean, standard deviation, minimum and maximum for each band was calculated to produce a 136-dimensional feature vector. The ASC, ASS and ASF were analysed to produce the mean, standard deviation, minimum and maximum giving a 4-dimensional feature vector for each.

Temporal Timbral descriptors describe the signal power function over time. The Log Attack Time (LAT) descriptor characterizes the attack of a sound (the time it takes for the signal to rise from silence to its maximum amplitude). This feature signifies the difference between a sudden and a smooth sound. The Temporal Centroid (TC) descriptor computes a time based centroid as the time average over the energy envelope of the signal. These features are not calculated on a frame-by-frame basis and so a multi-dimensional feature vector is not available, only the single feature.

4. CLASSIFICATION

Automatic emotion recognition in unaccompanied singing is a less well researched field than speech or music and so, in this work, the emotion range is limited to the discrimination between positive and negative emotions only. We are aware that the state-of-the-art in both speech and music communities is to discriminate between a far greater range of emotions, but we have yet to ascertain whether this is possible or useful for singing. A dataset comprising all songs from all singers (denoted as all in the results) was used for classification. We also present results from a single female singer to highlight the need for increased signal normalization prior to processing and the excellent results that might be attainable if we can successfully remove the singer bias.

For each input feature set, a four-fold cross validation is conducted with a Radial Basis Function kernel. This kernel is popular amongst the MIR community (Muyan, Naiyao & Hancheng, 2004; Han, Ho, Dannenberg & Hwang, 2009). The implementation used the scikit-learn (Pedregosa, Varoquaux, Gramfort, Michel & Thirion, 2011) Python module. The parameters for the SVM were (c=1.0, kernel=rbf, degree=3, gamma=0.0, coef=0.0, shrinking=true, tolerance=0.001, cache size=200, class weight=none). The SVM was run for 100 iterations with randomly split testing and training sets. The results presented in table 1 give the mean percentage of songs with were correctly classified according to emotion. Boxplots in figure 1 and figure 2 are also given to illustrate the overall classifier performance.

dataset	pitch (V)	shimmer (V)	jitter (V)
female	78	76	75
all	60	61	61
	audio power (VUV)	HNR (V)	MFCCs
female	74	74	76 (V)
all	60	61	73 (VUV)
	formants (V)		
	combined	burg	keep all	robust
female	77	75	77	75
all	61	66	61	61
	timbral descriptors (V)		
	combined	TC	LAT
female	72	73	76
all	60	60	57
	spectral descriptors (VUV)				
	combined	ASE	ASC	ASF	ASS
female	82	86	86	86	77
all	69	64	64	64	63

Table 1: SVM success rates for feature sets. The best performing features extracted from voiced frames only (V) and voiced and unvoiced frames (VUV) are presented. The spectral and timbral features are presented as individual features and as combined feature sets.

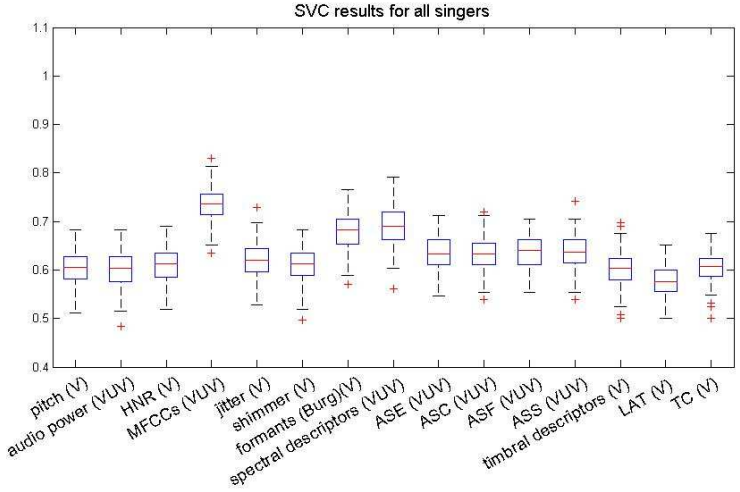


Figure 1: Boxplot showing the median, 25th and 75th percentiles, range and outliers for the SVM classification of the dataset containing all singers.

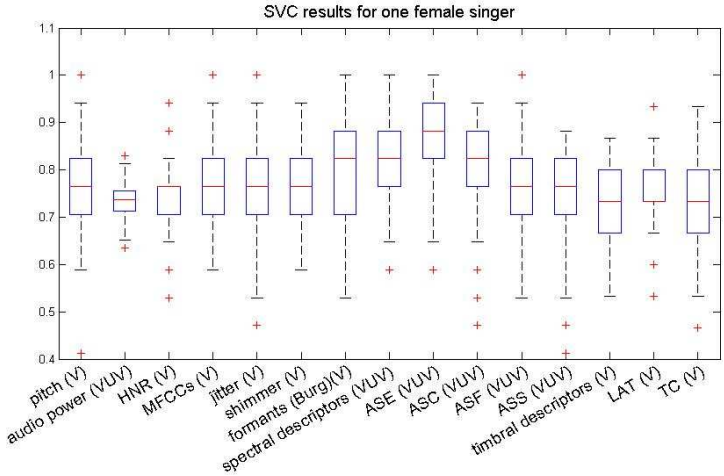


Figure 2: Boxplot showing the median, 25th and 75th percentiles, range and outliers for the SVM classification of the dataset containing one female singer.

4.1. Classification Results

The SVM results summarised in table 1 are the percentage of correctly identified songs. Bracketed terms indicate frame selection where (V) = voiced frames only, (VUV) = voiced and unvoiced frames. Results for V and VUV frames typically differed by a few percent and so only the best result is presented. Results for datasets containing audio from all singers (all) and one female singer are presented. Boxplots showing the SVC accuracy are also presented in figure 1 and figure 2 to give an indication of the SVM performance over 100 iterations.

5. DISCUSSION

Classification accuracy for the single singer is typically 10% or higher than that achieved for the data-set of all singers. All results are good, but the MFCCs, formants (Burg) and combined spectral descriptors are the best performers. We can conclude that, for singing, all these features can be considered strong emotional cues. Comparing boxplots we see that, although the median results for the female singer are better than for all samples, the classifier shows a greater variation in performance. This is likely due to the smaller dataset available for a single singer. Feature sets are discussed individually in the remainder of this section.

The classification rates for the HNR are disappointing. In speech the HNR is a clear indication of emotion, however, in singing this appears not to be the case. This might be explained when one considers that, in Chinese Opera (and possibly other modes of singing) unvoiced utterances are deliberately replaced with more voiced utterances as these are easier to sustain at high volume. Singing is hence identified as a special case different from speech.

The MFCCs perform well, but have been shown to also perform well for both speech and music emotion detection and so this not entirely surprising.

Both the timbral and spectral descriptors show good classification rates (as individual descriptors and combined). Timbral features are related, in speech, to articulation and hence clearly have direct relevance to the emotional state of the singer. The spectral properties characterise the overall frequency content and distribution of the audio, an approach shown to be successful in both music and speech. Their good performance confirms their usefulness when analysing singing.

The performance of the formant feature is also worth commenting on. Classification rates are high for both the single singer dataset and the all singer dataset. The conclusion is that singers transmit strong emotional content through the manipulation of their vocal formants. Informal discussion with singers supports this hypothesis. They describe the process of 'channeling their energy through their vocal cavity' and shaping that energy to deliver the desired emotions (Chapman, 2006). This is perhaps a conscious manipulation of the vocal tract to encode emotional content. A study of how formants are manipulated for encoding specific emotions would be a significant step towards designing useful audio analysis tools for singers.

There is an additional reason that formants may perform so well for emotion recognition in singing. In (Ververidis & Kotropoulos, 2006) it is explained that, for emotion recognition in speech, the formants are generally seen to perform poorly because there is a problem with correct identification of the formants. For example the second formant (F2) is often confused with the first formant (F1) and F1 can be confused with the fundamental frequency. This is not likely to occur for singing as the fundamental frequency will be clearly separated from the formants in the frequency domain due to the much higher pitch range employed.

6. CONCLUSIONS AND FURTHER WORK

This paper demonstrates that methods used to detect emotion in speech and music can be applied successfully to singing voice. Our hypothesis that, when it comes to MIR, singing should be viewed as the intersection of speech and music is well supported by our findings. Feature sets that have a good performance record when used to discriminate emotion in speech are shown to work either well for singing (pitch), poorly for singing (HNR) or to have a much improved performance for singing (formants). The change in feature performance is supported by singing theory.

The next step is to expand the feature set and the number of emotions detected in order to determine the best feature set for identifying emotional cues in singing voice. We acknowledge that our current dataset has a bias towards Chinese Opera and that not all emotional cues will be culturally independent. Further work will include a cross-cultural comparison including both western and non-western singing to support the hypothesis that emotional cues in singing are largely culture independent. Also of interest will be to perform feature extraction on singing that has been separated from a multi-track mix to see how source-separation algorithms effect feature extraction and resulting classifications.

7. REFERENCES

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In IFA Proceedings, 17, *pages 97–110.*

Boersma, P., & Weenink D. (2012). PRAAT: doing phonetics by computer. url=http://www.praat.org.

Chapman, J. L. (2006). Singing and Teaching Singing: A Holistic Approach to Classical Voice. Plural Publishing Inc., San Diego CA.

Davis, P. (1998). Emotional influences on singing. Australian Voice, *4:13–18.*

De Cheveigne, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. Journal of The Acoustical Society of America, *111.*

*Dong, L., Sundberg, J., & Kong, J. (2014). Loudness and pitch of Kunqu Opera.*Journal of Voice, 28*(1), 14-19.*

*Dong, L., Kong, J., & Sundberg, J. (2013). Long-term-average spectrum characteristics of Kunqu Opera singers' speaking, singing and stage speech.*Logopedics Phoniatrics Vocology, *(0), 1-9.*

El Ayadi, M., Kamel, M.S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition, *44(3): pages 572–587.*

Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Pedregosa, F., & Varoquaux, G. (2011). Scikit-learn: Machine learning in python. The Journal of Machine Learning Research, *vol. 12.*

France, D.J., Shiavi, R.G., Silverman, S., Silverman, M., & Wilkes, M. (2000). Biomedical Engineering, IEEE Transactions on, *47(7): pages 829–837.*

Fujihara, H., & Goto, M., (2007). A music information retrieval system based on singing voice timbre. Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR), *pages 467–470.*

Goto, M., Saitou, T., Nakano, T., & Fujihara, H., (2010). Singing information processing based on singing voice modelling. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), *pages 5506 – 5509.*

ISO/IEC international standard 15938. Information technology mpeg7 multimedia content description interface. part 4: Audio.

Han, B.J., Ho, S., Dannenberg, R.B., & Hwang, E. (2009). Smers: Music emotion recognition using support vector regression. Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR), *pages 651–656.*

Juslin, P.N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological bulletin, *pages 770–814.*

Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A., & Turnbull, D. (2010). Music emotion recognition: A state of the art review. Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR), *pages 255–266.*

Lampropoulos, A.S., & Tsihrintzis, G.A. (2012). Evaluation of mpeg-7 descriptors for speech emotional recognition. Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), *pages 98–101.*

Laurier, C., & Herrera, P. (2007). Audio music mood classification using support vector machine. Music Information Retrieval Evaluation eXchange (MIREX) *extended abstract.*

Li, X., Tao, J., Johnson, M.T., Soltis, J., Savage, A., Leong, K.M., & Newman, J.D. (2007). Stress and emotion classification using jitter and shimmer features. Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), *pages 1081–1084.*

Lu, L., Liu, D., & Zhang, H.J., (2006). Automatic mood detection and tracking of music audio signals. Audio, Speech, and Language Processing, IEEE Transactions on, *pages 5–18, Jan. 2006.*

Muyuan, W., Naiyao, Z., & Hancheng, Z. (2004). User-adaptive music emotion recognition. Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on *(Vol. 2, pp. 1352-1355). IEEE.*

Petrushin, V.A. (2000). Emotion recognition in speech signal: experimental study, development, and application. Proceedings of the sixth international conference on spoken language ICSLP, *vol 2 pages 222–225.*

Pohle, T., Pampalk, E., & Widmer, G. (2005). Evaluation of frequently used audio features for classification of music into perceptual categories. Proceedings of the International Workshop on Content-Based Multimedia Indexing.

Qu, C., & Liu Y. (2006). Acoustic analysis and characteristics of vocal range in Beijing Opera actors. url=http://www.ncbi.nlm.nih.gov/pubmed/12768694.

Rapoport, E. (1996). Emotional expression code in opera and lied singing. Journal of new music research, *25(2):109–149.*

Scherer, K.R. (2003). Vocal communication of emotion: A review of research paradigms. Speech communication, *40(1):227–256.*

Siegwart, H., & Scherer, K.R. (1995). Acoustic concomitants of emotional expression in operatic singing: the case of lucia in ardi gli incensi. Journal of Voice, *9:249–249.*

Sundberg, J., Iwarsson, J., & Hagegåard, H., (1995). A singer's expression of emotions in sung performance. Vocal fold physiology: Voice quality control, *pages 217–229.*

Sundberg, J., Gu, L., Huang, Q., & Huang, P. (2012). Acoustical study of classical Peking Opera singing. Journal of Voice, 26*(2), 137-143.*

Tsai, C.G., Wang, L.C., Wang, S.W., Shau, Y.W., Hsiao, T.Y., & Auhagen, W. (2010). Aggressiveness of the growl-like timbre: acoustic characteristics, musical implications, and biomechanical mechanisms. Music Perception, *27:209–221.*

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. Speech Communication, *48(9):1162 – 1181.*

Wang, L., Li, A., & Fang, Q. (2006). A method for decomposing and modelling jitter in expressive speech in Chinese. In Proceedings of the ISCA conference on Speech Prosody, *Dresden.*

Yeh, J., Pao, T., Lin, C., Tsai, T., & Chen, Y. (2010). Segment-based emotion recognition from continuous mandarin Chinese speech. Computers in Human Behavior, *pages 1545–1552.*

Zhang, Y., Zhou, J., & Wang, X. (2008). A study on Chinese traditional opera. In International Conference on Machine Learning and Cybernetics, *2008.*

SINGING INDOORS AND SINGING OUTDOORS: FORMANT TUNING

Rytis Ambrazevičius[1,2]
[1]*Kaunas University of Technology, Lithuania,* [2]*Lithuanian Academy of Music and Theatre, Lithuania*
Time: 11:30

1. Background: It is generally suggested that singing indoors (a concert hall and, especially, a small room) differs from the singing outdoors (in open air), in different qualities. However, this issue is not widely studied, at least, in the cross-cultural context. The author of the present paper has discussed one example of Lithuanian traditional rye-cutting song. It was shown that partial formant tuning (formant technique) is applied when singing outdoors. It means that the formants are partially tuned to the partials of voice spectrum to intensify the voices.

2. Aims: The paper aims to compare the examples of Lithuanian traditional singing representing the 'singing indoors' and 'singing outdoors' in different ethnic regions and genres, and to verify the presumption about the role of formant technique as the differential cue for the two modes of singing.

3. Method: Acoustical analysis of sound recordings (measurements of pitches, SPLs, and formants) and listening tests on perception of phonetics are applied. Statistical analysis is used for the generalizations.

4. Results: Set of 10 examples of the traditional open-air songs (mostly rye-cutting and hay-making songs) and the songs performed indoors was composed. The analysis shows that formant tuning is applied in the performances of the open-air songs, even if they were recorded in a sound studio. It is not characteristic of singing indoors. The difference in articulation is demonstrated by the listening test: if the linguistic context is removed, the differences between isolated prolonged vowels in the open-air songs are hardly perceived (the vowels become similar), whereas the differences are still well perceived in the case of singing indoors. Moreover, the formant tuning is characteristic mostly of the prolonged notes (i.e. anchor tones shaping the nucleus of the musical scale) in the open-air songs. Thus vocal technique assists in the construction and perception of the tonal hierarchy.

5. Conclusions: It can be stated that, based on the examples of Lithuanian traditional singing, 'singing indoors' and 'singing outdoors' differ substantially in the level of application of formant tuning. The tuning is applied clearly in the case of 'singing outdoors', thus making the vocal technique more 'economic' in terms of vocal efforts, more 'covered', and thus more similar to the Western academic singing, whereas 'singing indoors' is relatively more speech-like.

MAKING MYSELF UNDERSTOOD: FACTORS AFFECTING THE UNDERSTANDING OF SUNG TEXT

Philip Fine[1], Jane Ginsborg[2], Christopher Barlow[3]
[1]University of Buckingham, United Kingdom, [2]Royal Northern College of Music, United Kingdom, [3]Southampton Solent University, United Kingdom
Time: 12:00

ABSTRACT

Understanding sung lyrics is thought to be an important component of listeners' enjoyment of song, and of central concern to singers and, for instance, choral conductors. This paper describes an exploratory survey and two follow-up empirical studies which investigated the factors affecting sung text intelligibility. An exploratory survey of 143 musicians, including singers, singing teachers, and listeners, yielded 43 discrete factors perceived to affect sung text intelligibility. These were categorized as performer-related, listener-related, environment-related, and words/music-related. Most respondents felt it important to be able to understand the words, even in an unfamiliar language, although understanding text in a familiar language was considered more important. Based on the survey results, two listening studies were conducted. Number of singers, song familiarity, listener expertise, and meaningfulness of the text were manipulated, and participants wrote down all the words they could understand from sung renditions of 4-line unaccompanied songs. A soloist was easier to understand than a choir, and words were more intelligible when they had been heard before and were semantically meaningful. Singers were better at comprehending lyrics than non-singers.

1. INTRODUCTION

Singing is universal, and is unique in combining music and language in the form of song. One of the singer's foremost responsibilities is to communicate the sung text, and understanding the lyrics is assumed to be important for listener enjoyment and aesthetic appreciation (Omar Ali & Peynircioğlu, 2006), although this may be more true for some genres than others. Successful communication requires the listener to be able to understand at least some of the words being sung, and this relies on the singer singing clearly and intelligibly. Intelligibility is a measure of how understandable the singer's (or speaker's) message is to the listener (Kennedy & Trofimovich, 2008). Singing intelligibly is of particular concern for choirs and vocal ensembles, as well as, for instance, opera singers, and is a central aspect of vocal technique (e.g. Adams, 1998, Emmons & Chase, 2006): vocal pedagogy often emphasizes the development of good diction. However, the comprehension of sung text depends on multiple factors, as does the perception of speech (Harley, 2014), and only some of these factors relate to the performer. Others may relate to the listener, the environment, and the music and word-setting. In order to identify these factors, it seemed wise to ask musicians themselves what factors they believe affect sung text intelligibility, and indeed what importance they ascribe to being able to understand sung lyrics. To this end, an exploratory survey of musicians was undertaken.

Following this, four specific factors, which were revealed in the survey as being important in affecting sung text intelligibility, were manipulated in two empirical listening studies in order to provide experimental evidence to support the musicians' views. Much research into performer-related factors affecting sung text intelligibility (Scotto di Carlo 2007a, 2007b) has been carried out on isolated syllables (e.g. Sundberg, 1987) and individual words (e.g. Collister & Huron, 2008, Johnson, Huron & Collister, in press), so it was felt important to use more ecologically valid stimuli, in this case 4-line unaccompanied songs. The effect of four specific factors on listeners' comprehension of sung text was investigated, the four factors being: 1) the number of singers; 2) familiarity with the song; 3) meaningfulness of the song; and 4) musical experience of the listener.

To summarize: the first aim was to determine through an exploratory survey (Fine & Ginsborg, in review) the extent to which the intelligibility of sung text is important to a group of musicians and the factors they believe most likely to affect it, and the second aim was to determine in two experiments the effects on intelligibility of manipulating some of the factors most frequently identified by survey respondents.

2. SURVEY

2.1. Method

Musicians including singers and singing teachers, having provided demographic information, were asked to rate the importance of understanding sung text when listening to vocal and choral music in both familiar and unfamiliar languages. They were then invited to list those factors they believe affect sung text intelligibility, under four headings: performer-related, listener-related, environment-related, and words/music-related. There were 143 respondents, 61% female and 39% male, aged between 18 and 67 (mean=35.8, SD=13.7 years). More than three-quarters of the respondents (76%) were from the UK, the remainder from Europe, the Americas, and Australia. Over half (56%) were professional or semi-professional musicians or student singers, and were deemed experts: the remainder (44%) were amateur or occasional singers, or non-singers, and were deemed non-experts. Eighteen of the experts were or had been singing teachers with a mean of 10.2 years' teaching experience (SD=11.0 years).

2.2. Results

Most survey respondents felt it very or quite important to be able to understand sung text in both a familiar language (94%) and an unfamiliar language (68%), although understanding lyrics in a familiar language was deemed significantly more important ($p<.001$). Experts rated understanding lyrics in both familiar and unfamiliar languages as significantly more important than non-experts ($p<.025$).

Ninety-four respondents provided 394 open-ended statements in response to the request for factors affecting sung text intelligibility. These were split into 851 sub-statements, each referring to one factor, which were considered independently by two of the authors. Of these 851 sub-statements, 287 (34%) referred to the performer, 244 (29%) to the environment, 203 (24%) to the listener and just 117 (14%) to music and word-

setting. Following discussion over any categorization differences, 43 separate factors were extracted and categorized under the four headings suggested in the questionnaire. Non-experts listed significantly more factors than experts, both overall (p=.032) and listener-related (p=.004).

The four most commonly mentioned of the 15 performer-related factors, accounting for 57% of all performer-related sub-statements, were: articulation, diction, and enunciation; balance between singer(s) and accompaniment; communication of text, expression, and stage presence; and attitude, effort and projection. The four main environment-related factors mentioned, out of a total of 10, were acoustic, location, distraction, and the use of amplification or records. These accounted for 73% of the environment-related sub-statements.

There were 11 listener-related factors mentioned: the four most popular, accounting for 70% of the listener-related sub-statements, were hearing ability, attention, familiarity, and motivation. Finally, of the 7 factors related to the words and music, the most popular four were genre, the relationship between the words and music, compositional style, and language: these accounted for 76% of words/music-related sub-statements.

2.3. Discussion

Understanding sung text was generally felt to be important to musicians, even in unfamiliar languages, and certain factors were mentioned many times. However, the survey was necessarily subjective, and although musicians may perceive certain factors as important, this does not necessarily mean that they do in fact affect comprehension. Therefore, four of the factors identified in the survey were selected for further empirical investigation: number of singers (performer-related); meaningfulness of the text (music/words-related); familiarity with the song (listener-related); and listeners' singing experience (listener-related). It was hypothesized that fewer singers, greater singing experience (on the part of the listener), greater familiarity and more meaningful texts would all increase understanding.

3. EXPERIMENTAL STUDIES

3.1. Method

3.1.1. Participants

In Experiment 1 there were 48 participants (15M, 23F). Half were self-reported singers (age range 19-54, median 36 years), and half self-reported non-singers (age range 17-60, median 30 years). In Experiment 2 there were also 48 participants (25M, 23F), with a broader age range: 24 singers (age range 19-81, median 24.5 years) and 24 non-singers (age range 18-62, median 20 years).

3.1.2. Materials

In each experiment, songs consisting of 8-bar melodies were used as stimuli (as in Ginsborg, 2004). In Experiment 1, four songs were sung by a soprano soloist and by a trio of soprano, tenor and bass; all the singers were professionals or semi-professionals. Two were settings of short poems, and two were versions in which the semantically-meaningful words of the originals had been replaced with a sequence of numbers interspersed with short words ("and", "the", and "no").

Two additional songs were created for Experiment 2, in which non-semantically-meaningful, non-syntactical versions of all four songs were produced by scrambling the lyrics. They were all sung by a solo soprano and an unaccompanied chamber choir singing in unison. In both experiments, the stimuli were recorded at 48kHz, 24 bit resolution onto a digital audio workstation, using a Neumann KM130 omnidirectional condenser microphone placed a suitable distance from the singer(s), and then encoded as mp3 files at a constant bit rate of 320 kilobits per second. Stimuli were presented on a laptop at a comfortable volume using either its internal or external speakers.

3.1.3. Procedure

The procedure was identical in both experiments. Once the participants had provided demographic information about their singing experience and listening habits (in Experiment 2, participants also kept a listening and singing diary for a week before the listening session), they were provided with a piece of paper and two pens of different colors. The participant then heard each song played twice, with a pause after each line of text in the first playing. The participant was instructed to write down all the words they heard. The second time it was played through without stopping, and the participant was asked to use the other pen to indicate any alterations or addition to the heard words. The order of the songs and conditions were partially counterbalanced.

In Experiment 1, missing, additional, and incorrectly heard words were counted as errors, but incorrect word order was ignored as memory was not being tested. Errors were subtracted from the total possible number of words in the song, and a percentage performance score was calculated. The scoring technique was slightly refined for Experiment 2, using the informational semantic match (ISM) method (Hustad, 2006), developed for use with dysarthric speakers: individual syllables were scored rather than words, and inserted words were ignored. Again an overall percentage performance score was calculated. In both studies, two researchers scored the productions independently, and disagreements were resolved following discussion.

3.2. Results

Mixed ANOVAs were carried out on the mean performance data for Experiment 1 (see Table 1). Within-subject variables were number of singers (solo vs. trio), hearing (first vs. second), and text type (words vs. numbers); participants' singing experience was a between-subject factor. There were significant main effects of hearing ($F_{1,46}$=82.13, p<0.001), with better performance on second hearing, singing experience ($F_{1,46}$=5.95, p<0.02), singers scoring higher than non-singers, and text type ($F_{1,46}$=7.65, p=.008), with numbers (93.7%) easier to understand than words (91.2%), but no effect of number of singers (92.9% solo, 92.0% ensemble).

	Singers	Non-singers
Solo – first hearing	92.1	89.3
Solo – second hearing	96.4	93.9
Trio – first hearing	91.3	87.6
Trio – second hearing	96.6	92.5

Table 1: Sung text understanding performance in Experiment 1 (mean percentage data).

There were significant positive correlations between performance and self-rated singing experience (r_{48}=0.31, p<0.05), number of years' singing (r_{48}=0.33, p<0.05), and hours of singing per week (r_{48}=0.45, p<0.01), and a negative correlation between performance and hours spent listening to sung music per week (r_{48}=-0.31, p=0.05).

In Experiment 2, a linear mixed model analysis was carried out on the percentage performance data (i.e., syllables transcribed correctly), using the same independent variables as for Experiment 1. The effects of all four independent variables were significant and are shown in Table 2. Again, singers identified significantly more words than non-singers and all participants wrote down more on the second hearing. In contrast to the findings of Experiment 1, meaningful text was more intelligible than scrambled, and the choir was less intelligible than a solo singer.

		Mean %	$F_{44, 132}$
Participants	Singers	67.0 (20.9)	
	Non-singers	55.8 (23.8)	9.2**
Performance	First	56.9 (23.2)	
	Second	65.9 (22.1)	36.2***
Text	Meaningful	76.3 (17.7)	
	Scrambled	46.5 (17.4)	398.0***
Performer	Soloist	65.6 (22.0)	
	Choir	57.1 (23.3)	33.1***

*p < .05 **p < .01 ***p < .001

Table 2: Main effects in Experiment 2 (SDs in brackets).

There were also significant positive correlations between performance on the task and hours' singing per week (r_{48}= 0.32, p<.05), and hours' listening to sung music per week (rho_{48}=0.39, p<.01), and negative correlations between performance and age (r_{48}=-0.35, p<.05), and, for the singers, years' singing experience (r_{48}=-0.54, p<.01).

3.3. Discussion

The results of both studies were broadly in agreement with one another. A choir was more difficult to understand than a soloist, although a trio was not significantly harder than a solo singer. The relatively large number of singers in the choir (over 20) would be expected to muddy the acoustic signal. However, as the trio were all trained singers taught to enunciate as clearly as possible, this may have masked the effect of more singers. In both experiments, participants recorded more words on the second hearing, demonstrating the importance of familiarity. Similarly, in both studies, singers found the sung text more intelligible than did non-singers. This supports the conclusion that the sharing of a culture in which it is acceptable to distort vowels and consonants, as they are normally produced in speech, can compensate for the "unnaturalness" of sung modifications, and also for the use of vibrato. Finally, in Experiment 2, scrambling the text had a large detrimental impact on comprehension, supporting the importance of context and top-down processing on text comprehension. This was a better way of removing semantic meaning than merely replacing words with numbers. These were generally perceived without difficulty since they are formed from a limited vocabulary of labels: indeed, in Experiment 1, performance in the 'numbers' condition was significantly better than in the 'words' condition.

Comprehension performance correlated significantly with singing experience (self ratings, number of years of singing, and hours of practice per week) and weekly hours spent listening to sung music. However, the direction of some of these correlations differed between the two experiments. This might have been a function of the different participant samples in the two studies. Experiment 2 included a less homogeneous sample with a broader age range: while two-thirds of the singers were students, the remainder were considerably older members of an amateur choral society, but only a few of the non-listeners were recruited from members of staff in their 50s and 60s. This might also have explained the finding in Experiment 2 that performance deteriorated with age. Finally, it was noticeable that overall performance in Experiment 2 (61.4%) was generally worse than in Experiment 1 (92.5%). There are several possible explanations for this. The scoring criteria were stricter in Experiment 2, and two new songs were introduced in Experiment 2, which were perhaps harder to comprehend; furthermore the scrambled texts were much harder to understand than the original texts, while it was easier to understand the numbers in Experiment 1, suggesting that it was the nature of the stimuli that caused the apparent decrement in understanding. The greater age range in Experiment 2 might also have reduced overall performance.

4. GENERAL DISCUSSION

Many factors are believed by musicians to affect the intelligibility of sung text: beliefs supported empirically, at least for the four factors tested. These findings have potential utility for singers, singing teachers, choral directors and composers. Certain limitations in the survey and studies are worth considering. Retrospective estimates of time spent singing and listening to music are open to error, although the use of a diary in Experiment 2 no doubt helped increase the validity of such estimates. There is, however, the issue of active vs. passive listening: people do not always notice background music, including whether it is sung, making estimating listening time harder. Measuring singing experience is also not straightforward: some participants used to sing but no longer did so: should they be counted as singers or non-singers? Also, many people sing informally ('Happy birthday') but it is questionable whether that makes them singers. Perhaps the use of stricter inclusion criteria including the use of singing training would be wise in future studies.

These studies give rise to a number of new research questions. Only musicians were surveyed, the majority of whom were performers, and a group of non-musicians would quite possibly list fewer factors but emphasize different ones from those identified in the original survey with musicians. Genre was an important factor in the survey, and, in general, different genres differ in their intelligibility (Condit-Schultz & Huron, 2013). It would be worth investigating the effect of genre on both the importance of understanding sung text and genre-specific factors influencing intelligibility in more detail. Finally, there are many similarities and differences between music and language (Patel, 2010) and also between speech and song, so it would be worth repeating the empirical study with stimuli spoken by both a solo speaker and a chorus of speakers. Future research will address some of these suggestions and explore some of the other factors identified in the survey.

5. REFERENCES

Adams, D. (1998). A handbook of diction for singers: Italian, German, French. Oxford: OUP.

Collister, L.B., & Huron, D. (2008). Comparison of word intelligibility in spoken and sung phrases. Empirical Musicology Review, 3 (3), 109-125.
Condit-Schultz, N., & Huron, D. (2013). Catching the lyrics: The intelligibility of lyrics in twelve genres. Paper presented at the SMPC, Toronto, Canada.
Emmons, S., & Chase, C. (2006). Prescriptions for Choral Excellence: Tone, text, dynamic leadership. New York: Oxford University Press.
Fine, P.A., & Ginsborg, J. (in review). Making myself understood: perceived factors affecting the intelligibility of sung text. Frontiers in Psychology.
Ginsborg, J. (2004). Singing by heart: memorisation strategies for the words and music of songs. In J. W. Davidson (ed.) The music practitioner: exploring practices and research in the development of the expert music performer, teacher and listener (pp. 149-60). Ashgate Press.
Harley, T.A. (2014). The Psychology of Language: From Data to Theory. 4th ed. Hove, England: Psychology Press.
Hustad, K.C. (2006). A closer look at transcription intelligibility for speakers with dysarthria: evaluation of scoring paradigms and linguistic errors made by listeners. American Journal of Speech-Language Pathology, 15(3), 268-277.
Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. Canadian Modern Language Review 64, 459-490.
Johnson, R., Huron, D., & Collister, L.B. (in press). Music and lyrics interactions and their influence on recognition of sung words: An investigation of word frequency, rhyme, metric stress, vocal timbre, melisma, and repetition priming. Empirical Musicology Review.
Omar Ali, S., & Peynircioğlu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners? Psychology of Music, 34(4), 511-534.
Patel, A. D. (2010). Music, Language, and the Brain. Oxford University Press, USA.
Scotto Di Carlo, N. (2007a). Effect of multifactorial constraints on opera-singing intelligibility (I). Journal of Singing 63 (4): 1–13.
Scotto Di Carlo, N. (2007b). Effect of multifactorial constraints on intelligibility of opera singing (II). Journal of Singing 63 (5): 559–567.
Sundberg, J. (1987). The Science of the Singing Voice. Dekalb, Illinois, USA: Northern Illinois University Press.

Session 11

[11A] Sadness in Music

SEOUL 14:00-15:00 Friday 8 Aug 2014

WHY LISTEN TO SAD MUSIC? EXPLORING THE UNDERLYING REASONS

Henna-Riikka Peltola[1], Jonna Vuoskoski[2], Tuomas Eerola[3]
[1]Finnish Centre for Interdiciplinary Music Research, Department of Music, University of Jyväskylä, Finland, [2]Faculty of Music, University of Oxford, United Kingdom, [3]Department of Music, Durham University, United Kingdom

Time: 14:00

In psychology, sadness is commonly considered a negative emotion that people tend to avoid. In the context of the arts, however, people have drawn enjoyment and entertainment from sadness and tragedy since the days of the ancient Greeks. Previous studies have shown that sad music is sometimes used for mood-regulation purposes, and that certain personality traits are connected with liking for sad music. However, little is still known about the variety of reasons and motivations for listening to sad music. The aim of the present study is to investigate people's attitudes towards sad music: What kinds of motives do people have for listening to sad music, and to what extent do they find it pleasurable? Furthermore, can any broader underlying themes be found in their responses? A survey was administered to a nationally representative sample of Finns (N = 386). Participants rated their agreement with 30 statements concerning attitudes and everyday uses of sad music. The statements were derived from a previous qualitative study that explored people's motives for listening to sad music. Participants' ratings were subjected to factor analysis, resulting in six factors explaining 51% of the variance (RMSEA = 0.049). The factors were labeled AVOIDANCE, MEMORY, APPROACH, APPRECIATION, SOCIAL, and AMPLIFICATION. Age and gender were associated with the factor scores: younger participants gave higher values for APPROACH, MEMORY, and AMPLIFICATION, while women gave higher values for MEMORY. The results of the survey suggest that listening to sad music elicits a wide variety of responses. Although nearly half of the participants tended to avoid sad music, the prevalence of the more positive categories nevertheless indicates that many listeners are able to draw comfort (APPROACH), perspective to life (APPRECIATION), and feelings of not being alone (SOCIAL) from listening to sad music. Some listeners even sought to intensify negative feelings (AMPLIFICATION) via listening to sad music, demonstrating the involvement of multiple and complementary strategies.

SAD LYRICS ARE SUNG SLOWER: BUT IS THIS A CONSEQUENCE OF VALENCE OR AROUSAL?

Kirsten Nisula, Evan Copeland, David Huron
School of Music, Ohio State University

Time: 14:30

ABSTRACT

In his detailed clinical analysis of depressed patients, Emil Kraepelin (1899/1921) observed that sad speech is characterized by a slower speaking rate. Various studies have also noted parallels in "sad" music (Juslin & Laukka, 2003; Post & Huron, 2009). Sadness is a low arousal-negatively valence affect (Russell, 1980), however, low-arousal alone is associated with decreased levels of epinephrine and acetylcholine known to effect speaking rate (Huron, 2012). Is the association between slow tempo or speaking rate and sadness really a consequence of sadness, or could it simply be a consequence of low arousal? In this paper, we present three studies that explore the conjecture that slower speaking rates will be found in music and lyrics expressing low-arousal rather than simply "sadness."

For all three studies, our dependent measure is syllable rate (syllables/second) calculated from annotated audio recordings. In the first study, we compare the syllable rate of 25 randomly selected Billboard Top 40 songs deemed to be in the minor mode (Schellenberg, 2013) with 25 major-mode songs from matched years. Contrary to our expectation, we found that minor-mode songs exhibit a faster syllable rate. However, in our second study, we compared the syllable rate of songs with lyrics rated by the authors as expressing "sadness" or "happiness." In this case we found the predicted association of sad lyrics with slower average syllable rate Finally, a third study examined the relationship between the arousal

and valence expressed by the lyrics and the syllable rate of the sung song. Results suggest that valence may be more important than arousal, although the two factors share considerable variance.

1. INTRODUCTION

Among the detailed clinical observations of his sad and depressed patients, Kraepelin (1899/1921) noted that sad speech is characterized by a slower speaking rate. Subsequent empirical studies are consistent with this observation. Breitenstein, Van Lancker, & Daum (2001), for example, showed that slower speech rate is associated with sadness in samples of both German and American English. Siegman and Boyle (1993) similarly found that slow speech rate is associated with sadness and depression.

A number of studies have noted parallels between prosodic cues and musical organization (Juslin & Laukka, 2003). Post and Huron (2009) for example, found that minor-mode works tend to be associated with slower notated tempos. An exception to this pattern is evident in 19th century music. Although slower tempo markings are significantly more likely to be associated with the minor mode for Baroque and Classical music, the reverse trend is observed in music from the Romantic period. In the same study, Post and Huron also analyzed audio recordings for 21 theme-and-variation keyboard works. They compared the performed tempos for neighboring variations that change mode. They found that performers tend to slow down when moving from a major-mode to minor-mode variation, and conversely, tend to speed up when moving from a minor-mode to major-mode variation.

For Western-enculturated listeners, the association of the minor mode with sadness has been observed for hundreds of years. Moreover, controlled empirical studies have also underscored this association (e.g. Hevner, 1935). However, the association is not ironclad. There are major-mode works that are widely deemed to be sad (e.g., the aria "He Was Despised" from Handel's *Messiah*), and minor-mode works that are widely deemed not to be sad (e.g., the opening section of Mozart's *Rondo Alla Turca*). Nevertheless, there appears to be a broad association between the minor mode and sadness–at least for Western-enculturated listeners.

The initial motivation for this research is to test the following hypothesis: Songs deemed to exhibit a sad affect tend to exhibit slower lyrics. In order to address this hypothesis, three studies were carried out, each employing a different sampling strategy. In all three studies, however, the music samples are restricted to English-language songs.

The first study examines the relationship between the minor mode and the speed of lyrics. The second study more directly tests the conjecture using lyrics from nominally sad songs. The third study addresses whether this relationship is simply due to low physiological arousal, and so contrasts sad lyrics with lyrics for songs deemed to represent *relaxed* or *contented* states.

2. STUDY #1

This study was motivated in part by Hevner's (1935) findings that minor-mode works are associated with sadness in the Western classical art music tradition.

2.1. Method

H1: Songs ending in the minor mode exhibit slower syllable rates than songs ending in the major mode.

Sampling Strategy. For this study, the songs were selected from a database assembled by Glenn Schellenberg containing the Billboard Magazine Top-40 songs for each year from 1965 to 2009 (Schellenberg, 2013). Schellenberg and his colleagues characterized the final modality for each song. They identified 693 songs exhibiting a major final modality and 307 songs exhibiting a minor final modality. Nine songs were deemed to have an ambiguous or undefined modality.

From Schellenberg's list of 1,009 songs, twenty-five songs were sampled using a random number generator. These selected songs were of either major or minor modality. A matched pair of the opposite modality was then selected for each song by proceeding down the list until another song from the same year with opposite modality was encountered. This sampling method resulted in twenty-five matched pairs consisting of one major modality song and one minor modality song for a total of fifty songs. Recordings were assembled for all 50 songs. For each recording, song lyrics were collected using the web.

Syllable rate calculation. For each of the sampled songs, lyrics were obtained from websites such as lyrics.com, azlyrics.com, metrolyrics.com, and others. The lyrics were checked against each respective sound recording in order to ensure accuracy. In particular, it is important to check that all performed repetitions are present in the transcribed lyrics. Where necessary, the lyrics were modified to bring them into agreement with the actual sung material.

After establishing the accuracy of the lyrics, the number of syllables in each song was tallied using a specially designed computer program which provided syllable counts. Syllable counts for uncommon words or abbreviations found in the lyrics were identified by hand. Using this method a total syllable count was determined for each song.

Apart from the syllable counts, the total duration of singing was determined for each sound recording. This procedure was done using a computer stop-watch. The timer was initiated each time the principal vocalist is present and terminated each time the vocalist stops singing. In many recordings, songs included backup vocals. Usually, these vocals simply duplicated concurrent lyrics sung by the principal vocalist. In other cases, the backup vocalists repeated (with delay) or added other material independent of the principal vocalist. In these cases, the "duration of singing" depended on the transcribed lyrics. If the transcribed lyrics included material sung by backup vocalists in the absence of the principal vocalist, then these sung passages were included in the timing measures. In short, the total duration of singing was measured so that it was consistent with the transcribed lyrics. In determining the duration of singing, three independence measures were made, and the average value of the three measurements was used. Consequently, the complete data set consisted of a tally of the number of syllables and the duration-of-singing for each of the 50 target songs.

2.2. Results

The results for Study #1 are illustrated in Figure 1 showing the average syllable rate (per second) for the major and minor songs. Contrary to our hypothesis, songs ending in the minor mode were significantly faster than songs ending in the major mode ($t(24)= 2.8208$, $p < .01$).

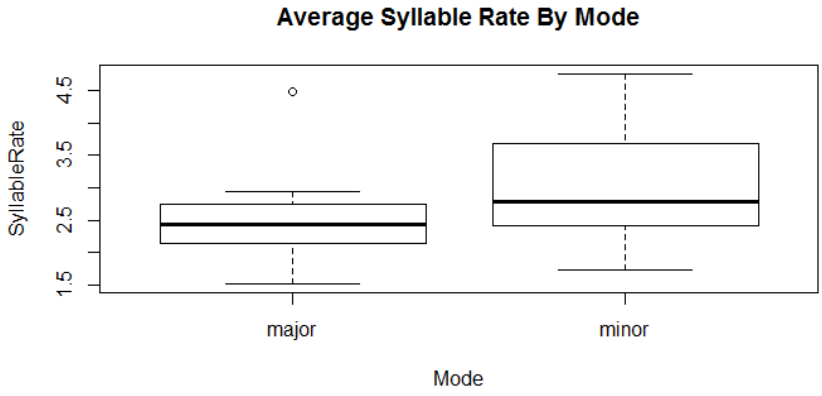


Figure 1. This boxplot displays the average syllable rate (in syllables per second) for songs ending in the major mode and songs ending in the minor mode.

3. STUDY #2

As noted, Study #1 relies on the assumption that there is an association between the minor mode and sadness. However, this relationship is not set. There are many minor-mode works that do not necessarily evoke or convey a sad affect, as in many works from the Romantic period (Horn & Huron, 2012). Consequently, it would be appropriate to test the hypothesis more directly by contrasting the syllable rates of songs whose lyrics are directly assessed to be "happy" or "sad," respectively.

3.1. Method

H2: Songs with lyric content judged to be "sad" exhibit slower syllable rates than songs with lyric content judged to be "happy."

Sampling Strategy. Lyrics were obtained via the website, azlyrics.com. This website included complete listings of English-language song lyrics and also provided search tools for accessing various songs. An initial search for candidate songs was completed by searching for the terms "sad," "sadness," "happy," and "happiness" in the song's title or lyrics. Songs by well-known artists or any songs known to the authors were excluded from the sample in order to reduce the possibility of familiarity influencing the judgment of the emotional content of the lyrics. Twenty complete song lyrics were randomly selected for each of the "sad" and "happy" search conditions.

Rating Procedure. Two of the authors read each of the sad and happy lyrics and independently rated the overall affective content of the lyrics using separate sadness and happiness scales. The independent ratings were averaged, producing a single score for each song lyric. Lyrics from the 10 highest scores for sad songs and the 10 highest scores for happy songs were selected for further analysis. Specifically, the syllable rate for each song was determined using the same method as described in Study #1.

3.2. Results

We compared the syllable rates for the sad and happy samples using a Welch two-samples t-test. We found that the syllable rates for the sad lyrics were significantly slower than the syllable rate for the happy lyrics (t=-2.9435, df=12.972, p=0.005718). In short, the results are consistent with our *a priori* hypothesis.

4. STUDY #3

It is common to characterize affect using a two-dimensional model of arousal and valence (e.g., Hevner, 1936; Russell, 1980). This model distinguishes four quadrants: high-energy/positive-valence affects (e.g,. joy), low-energy/positive-valence affects (e.g., contentment), high-energy/negative-valence affects (e.g., anger), and low-energy/negative-valence affects (e.g., sadness). Since we found an association between sadness and slow syllable rates for lyrics in Study #2, we might consider whether this relationship arises simply because sadness is associated with low physiological arousal. That is, we might ask whether negative valence is a necessary condition for slow lyric rate. Specifically, can we observe a low syllable rate for any affect that is linked to low arousal (such as relaxation, contentment, comfort, peace, etc.)? Accordingly, it would be helpful to contrast syllable rates for songs exhibiting nominally sad lyrics, with syllable rates for songs exhibiting a low arousal/positively valenced affect (e.g., relaxed). Additionally, it would be appropriate to compare both positively- and negatively-valenced affects exhibiting high arousal.

4.1. Method

H3: The speed of lyrics is attributable only to the level of arousal, and is not attributable to the valence of the expression. That is, arousal accounts for a greater variance in syllable rate than either the "sadness" of the expression, or the "happiness" of the expression.

Participants. In order to better characterize the affective content of the lyrics, twelve participants were recruited from the Ohio State School of Music subject pool (7 female, 20.2 average years).

Sampling Strategy. Songs were selected according to the four quadrants of the valence × arousal model of emotion. Once again, lyrics were obtained from the website azlyrics.com. For this study, the aim was to select songs representative of the four quadrants in the valence × arousal model. For the high-energy/positive-valence affects, the search terms consisted of "joy," "thrill," "delight/delightful," "rapture," and "dance/dancing." For the low-energy/positive-valence affects, the search terms consisted of "peace," "relax," "content," "calm," and "serene." For the high-energy/negative-valence affects, the search terms consisted of "angry," "mad," "furious," and "rage." For the low-energy/negative-valence affects, the search terms consisted of "sad," "depress," "poor," "gloom," and "sorry." From these four categories, 68 lyrics were

randomly selected–16 for each of the four target states plus an additional song from each category to determine between-subject reliability. Once again, songs by well-known artists as well as any song known to the researchers were excluded in order to reduce bias.

Design. Each of the twelve participants examined 20 complete song lyrics, and rated each lyric according to four dimensions: *energy, negativeness, sadness*, and *happiness*. The first two scales (energy/negativeness) were intended to characterize the arousal and valence, respectively, of each song.

In order to establish within-subject reliability, 4 of the 20 lyrics judged by each participant were repeated at the end of the experiment. In addition, each participant judged four lyrics in common with all the other participants—permitting a test of between-subject reliability.

4.2. Results

For each of the 64 song lyrics, three independent judgments were made for each of the four variables of interest. However, due to time limitations, the results reported here are only for 20 of the songs. The syllable rates for the corresponding recordings were calculated using the same procedure described in Study #1 by two of the authors. For each song, there are five measures: (1) the lyric syllable rate, (2) the average energy rating for the lyrical content, (3) the average negativity rating for the lyrical content, (4) the average sadness rating for the lyrical content, and (5) the average happiness rating for the lyrical content.

Testing the hypothesis from our first and second studies, we used multiple regression analysis to predict the syllable rate based on the combination of happiness and sadness ratings. The regression equation is consistent with syllable rate increasing with happiness ratings and decreasing with sadness ratings. While the relationship is significant (p=0.04), the effect size is small, with an adjusted R-squared of 0.23.

However, the main question posed in this study relates to shared variance and explanatory power for arousal and valence. When predicting syllable rate, the correlation between arousal (energy) estimates and syllable rate was found to +0.496. By comparison, the correlation between valence (negativity) estimates and syllable rate was +0.506. (That is, more negatively valenced lyrics were associated with faster syllable rates.) In order to determine shared variance, we used both energy and negativity estimates as predictors for syllable rate. Together, a significant linear model was observed, with an adjusted R-squared of 0.594 ($F(2,17)=14.88$, p=0.0002). This suggests that roughly 85 percent of the variance is shared in common between energy and negativity. Yet, despite the shared variance, the correlation between energy and negativity proved surprisingly low (-0.21). Given the high correlation between valence and syllable rate, the results are not consistent with the suggestion that arousal alone accounts for slow syllable rate. That is, valence plays a formative role.

5. GENERAL DISCUSSION

In the three studies reported here, we have investigated the relationship between syllable rate in sung lyrics and various other measures. In Study #1, we tested the prediction that syllable rate would be sung slower in nominally minor-mode works compared with nominally major-mode works. The underlying assumption is the widespread association of the minor mode with sadness for Western-enculturated listeners. However, contrary to our hypothesis, minor-mode works exhibit a statistically significant faster syllable rate.

In Study #2, we tested more directly the relationship between sadness and syllable rate by ignoring modality, and relying on judgments of sadness/happiness for specific song lyrics. In this study, we found the anticipated relationship between nominally sad songs and slower sung lyrics. In light of the results for Study #2, it appears that the results of Study #1 are suggestive of a changing use of the minor mode in contemporary popular song. That is, it appears that the minor mode in this repertoire has no particular affinity for the expression of sadness. As Post and Huron (2009) found in the 19th century, the minor mode appears to have a stronger association with fast tempos (or at least fast lyrics). This relationship suggests further stylistic studies of the use of major and minor modes in popular song is warranted.

In Study #3, we focused on the possibility that sung syllable rate may be related solely to arousal (energy level), and that valence (negativity) plays little or no role in lyric tempo. In this study we examined the relationship between syllable rate and independent ratings of arousal and valence for each song. Although both energy and negativity exhibited high correlations with syllable rate, valence (negativity) accounted for slightly more of the variance (although perhaps not significantly more). Furthermore, a multiple regression analysis suggests that, despite significant shared variance, both energy (arousal) and negativity (valence) make statistically significant contributions to predicting syllable rate.

Since only a subset of our total data was available for this analysis, we look forward to further analyses of the data related to Study #3.

6. REFERENCES

Breitenstein, C., Van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. Cognition and Emotion, 15*(1), 57–79. doi:10.1080/0269993004200114*

Hevner, K. (1935). The affective character of the major and minor modes in music. American Journal of Psychology1, 47, *103–118.*

Hevner, K. (1936). Experimental studies of the elements of expression in music. American Journal of Psychology, 48, *246–268.*

Horn, K., & Huron, D. (2012). Major and minor: An empirical study of the transition between Classicism and Romanticism. In E. Cambouropoulos, C. Tsougras, P. Mavromatic, & K. Pastiadis (eds.), 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music *(pp. 456–464). Thessaloniki.*

Huron, D. (2012). Understanding music-related emotion: Lessons from ethology. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiadis (eds.), Proceedings of the 12th International Conference on Music perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music *(pp. 473–481). Thessaloniki.*

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? Psychological Bulletin, 129*(5), 770–814. doi:10.1037/0033-2909.129.5.770*

Kraepelin, E. (1921). Psychiatrie. Ein Lehrbuch fur Studierende und Artz, ed. 2. *(R. M. (translator) Barclay, Ed.). Edinburgh: E.& S. Livingstone.*

Post, O., & Huron, D. (2009). Music in minor modes is slower (Except in the Romantic Period). Empirical Musicology Review, 4*(1), 1–9.*

Russell, J. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39*(6), 1161–1178.*

Schellenberg, G. (2013). Private communication: Lis of Billboard *Top 40 Final Modality.*

Siegman, A. W., & Boyle, S. (1993). Voices of fear and anxiety and sadness and depression: the effects of speech rate and loudness on fear and anxiety and sadness and depression. Journal of Abnormal Psychology, 102*(3), 430–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8408955*

[11B] Music in Digital Age

KYOTO 14:00-16:00 Friday 8 Aug 2014

BIG DATA FOR THE MUSIC PERCEPTION AND COGNITION COMMUNITY

Ichiro Fujinaga[1], David Sears[1], Andrew Hankinson[1]
[1]McGill University, Canada

Time: 14:00

Big Data is an umbrella term that describes datasets so large and complex that it requires special tools to collect and examine this data. This is opening up new avenues of research and experimentation in the music perception and cognition community (for example, Music Perception is devoting two special issues to musical corpus research in 2013-2015). Until now we have lacked the necessary empirical tools to rigorously document and explore the listening habits of everyday listeners. This presentation will provide the music perception and cognition researcher with an arsenal of useful tools for studying listeners (and the musical corpora to which they are exposed) on an unprecedented scale. The publicly available databases we will discuss include: Searchable databases: Liber Usualis (http://ddmal.music.mcgill.ca /liber/): A searchable database for the 2,340 pages of Liber Usualis (a book of Gregorian chants) with over 3 million n-grams. SIMSSA (Single Interface for Music Score Searching and Analysis) (http://simssa.ca): An international project of converting digitized music scores to searchable database music as full text. Music analysis databases: Harmony: The Billboard Dataset (http://ddmal.music.mcgill.ca/billboard): A database of harmonic analysis of 1,000 popular songs from the Billboard "Hot 100" chart from 1958 to 1991. The database contains over 400,000 labeled beats. Form: SALAMI: Manually annotated 1,000 pieces of music (audio) for training machine-learning algorithms for automatic structural analysis. We are in the process of analyzing approximately 350,000 pieces of audio. Vertical intervals: ELVIS (Electronic Locator of Vertical Interval Successions) (http://elvisproject.ca): A collection of searchable symbolic representation of European music from 16th to 19th century. It represents music from 100 composers and over 6,000 pieces / movements with approximately 3 million notes.Listeners database: The Million Listener's Database: A collection of real-time listening records (the time stamp and the name of the song) of approximately one million listeners from Last.fm.In addition to these we will also introduce several third-party music data services that may be used to enhance, support, or validate music research, including Echonest, The Million Song Database, and Peachnote.

MOOD EXPRESSION IN REAL-TIME COMPUTER GENERATED MUSIC USING PURE DATA

Marco Scirea[1], Yun-Gyung Cheong[1], Byung-Chull Bae[1]
[1]IT University of Copenhagen, Denmark

Time: 14:30

ABSTRACT

This paper presents an empirical study that investigated if procedurally generated music based on a set of musical features can elicit a target mood in the music listener. Drawn from the two-dimensional affect model proposed by Russell, the musical features that we have chosen to express moods are *intensity, timbre, rhythm,* and *dissonances*. The eight types of mood investigated in this study are being *bored, content, happy, miserable, tired, fearful, peaceful, and alarmed*. We created 8 short music clips using PD (Pure Data) programming language, each of them represents a particular mood. We carried out a pilot study and present a preliminary result.

1. INTRODUCTION

In this paper we present our empirical study to explore how the manipulation of a set of musical features can express different moods in music. As an affect model, we have employed the two-dimensional affect model proposed by Russell [16]: pleasant-unpleasant and arousal-sleepy. In addition to Russell's bipolar dimensional model, we also considered Thayer's explication for the arousal - including both energetic arousal and tense arousal [19]. The musical features that we employ to express moods are *intensity* (general volume), *timbre* (brightness), *rhythm* (strength, regularity, and tempo), and *dissonances*. The first three features (i.e., intensity, timbre, rhythm) were inspired by Liu's study [12] about mood information extraction from classical music pieces. We applied the same principles to generate music instead.

Moods are different from emotions; as defined by Beedie *et al.* [3], emotions are short and intense states, while moods are less strong and may last for much longer. Although music can evoke emotions in the listeners, we concentrated on moods in this paper, with the purpose to use this work for digital games. In games, we believe that gamers are likely to listen to the background music (e.g., background music) longer than music that would evoke a particular emotion.

The research question we aim to answer in this study was "can the music generated based on the combination of the features arouse different set of moods with a fine-granularity?" We selected eight types of mood for the study: *bored, content, happy, miserable, tired, fearful, peaceful,* and *alarmed*. To this end, we first located, based on Russell's study, the coordinates of the emotion words on the two-dimensional affect space. Then, we created eight clips of mood music according to the pleasantness-arousal coordinates obtained. To generate music for our study we employed a real-time procedural music generator that we had developed using PD (Pure Data) programming language [17]. Our music generation approach does not take into account chord sequences, leitmotifs, or improvisation. Instead, we aim to create a very minimalistic ambient music created by simple random number generators. This allows us to test our hypothesis of being able to display moods through only the manipulation of the mood defining features we consider. Despite a variety of studies towards procedural music generation [4], [6], [18], there has been little research on investigating the relationship between music and affect.

We carried out a pilot study that ten undergraduate/graduate students participated in. While listening to the generated 30 seconds music clip, each study participant specified the degree of pleasantness and arousal that they felt in the music, and then chose at most two mood words that could best represent the music. According to our early study results so far, only three affect words (*peaceful, alarmed,* and *bored*) were correctly perceived with reasonably good accuracies. Overall, however, the pleasantness/arousal answers from the participants were closer to our expected values.

2. MUSIC MOOD TAXONOMY

The set of adjectives that describe music mood and emotional response is immense and there is no accepted standard; for example Katayose *et al.* [7] use a set of adjectives including *Gloomy*, *Serious*, *Pathetic* and *Urbane*.

Russell [16] proposed a model of affect based on two bipolar dimensions: *pleasant-unpleasant* and *arousal-sleepy*, theorizing that each affect word can be mapped in this bi-dimensional space by a combination of these two components.

Thayer [19] applied Russel's model to music using as dimensions *stress* and *arousal* (see Figure 1). Although the name of the dimensions is slightly different from Russell's, their semantic meaning is the same. Since valence and arousal are commonly used terms, we will use these terms in this paper.

Thus the music is divided in four clusters as in Figure 1: **Anxious/Frantic** (Low Valence, High Arousal), **Depression** (Low Valence, Low Arousal), **Contentment** (High Valence, Low Arousal) and **Exuberance** (High Valence, High Arousal). These four clusters have the advantage of being explicit and discriminable. Also they are the basic music-induced moods (even if with different names) as discovered by Kreutz [9] (*Happiness*, *Sadness*, *Desire* and *Unrest*) and Lindstrom [11] (*Joy*, *Sadness*, *Anxiety* and *Calm*).

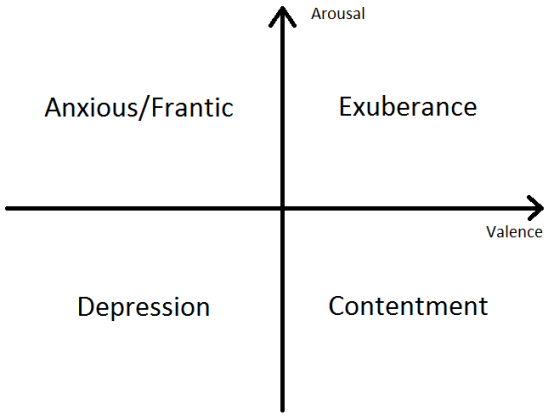


Figure 1: Thayer's bi-dimensional model [19].

3. MUSICAL MOOD FEATURES

In order to generate mood-based music, we used four musical features – intensity, timbre, rhythm, and dissonances, which are mainly inspired by Liu *et al*. [12]. While Liu et al.'s research focused on mood information extraction, we applied their approaches to generate music instead. This section extends our previous approach [17], introducing a new feature called dissonances.

3.1. Intensity

Intensity is defined by how strong the volume of the music is. It is an *arousal*-dependent feature: high arousal corresponds to high intensity; low arousal to low intensity.

Intuitively the more stress is present in the music, the more it will have a high volume. Calm pieces of music, in a similar manner, have a lower one.

3.2. Timbre

Timbre is what we could call the brightness of the music, that is, how much of the audio signal is composed by bass frequencies. It is often associated with "how pleasing to listeners"[1]. In previous literature audio features such as MFCC (Mel-Frequency Cepstral Coefficients) and spectral shape features have been used to analyze this timbral feature.

We associated this timbral feature with *valence*: the more positive the valence, the higher will the timbre be. The brightness of *Exuberance* music, for example, is generally higher than that of music in *Depression*, which will result in greater spectral energy in the high sub bands for *Exuberance*.

Generally, timbre is a factor that is very dependent on the instrumentation choice. In our case we acted on the synthesizers, our instruments, to generate brighter and darker sounds. In our generator we had three different sets of "instruments" (they were actually the same synthesizers with different settings to make them sound different) for high, low and neutral valence.

3.3. Rhythm

We included three features related to rhythm: strength, regularity and tempo [12].

- Rhythm strength: how prominent the rhythmic section is (drums and bass). This feature is *arousal* dependent.
- Regularity: how steady the rhythm is. This feature is *valence* dependent.
- Tempo: how fast the rhythm is. This feature is *arousal* dependent.

In a high valence/high arousal piece of music, for instance, we can observe that the rhythm is strong and steady. In a low valence/low arousal, on the other hand, the tempo is slow and the rhythm cannot be as easily recognized. We acted on these features in different ways. To influence rhythm strength, we changed how much the drums are prominent in the music. Having the instruments play notes on the beat or the upbeat created different feeling of regularity and irregularity. For example, in Contentment music, we favored a steady rhythm with notes falling on the

beats of the measure. In Depression music, on the other hand, we gave more space to upbeat notes. Finally, to influence the tempo we just acted on the BPMs (Beats Per Minute) of the music.

3.4. Dissonances

What we mean by dissonance is the juxtaposition of two notes very close to each other: for example C and C#. The distance between these two is just a semitone, which gives the listener a generally unpleasant sensation (like when you hear someone singing out of tune).

Dissonance doesn't mean that it always sounds poorly. In fact most music pieces contain dissonances, as they can be used as clues that express something's wrong. The listener's ear can also be trained to accept dissonances through repetition. In general, the bigger the interval between the two dissonant notes, the easier it is on the listener's ear: a C and a C# are always dissonant, but the dissonance is more evident if the notes are played from the same octave and not on two different ones.

Already in the first study we noticed that these features, originally devised to extract mood information, were enough to generate different moods. But we also realized that we could strengthen the impression by introducing dissonances in the music: for *Exuberance* and *Contentment* we use a diatonic scale, while for *Anxious* and *Depression* an altered one. We believe this is an important feature that cannot be ignored when wanting to show more precise moods in music.

Dissonance feature is *valence* depending. In our study we just used two scales: a *C major* scale (*C D E F G A B*) for positive and a *Eb Harmonic Minor* scale minus the third grade (*Eb F [Gb] Ab Bb B D*) for negative valence. Music built on a minor scales is generally considered more somber than when made in a major key. This is not technically correct in our system because it would require a grade of organization and harmony that would make plain which is the root note. The notes of the harmonic minor scale are the same as the natural minor except that the seventh degree is raised by one semitone, making an augmented second between the sixth and seventh degrees. For our unstructured music this means that we have a whole-and-a-half interval between B and D and two half intervals (*D-Eb* and *Bb-B*). The removal of the third grade (*Gb*) makes even more difficult to the listener's ear to identify the key, effectively making the dissonances sound as such.

4. AN EXPERIMENTAL PILOT STUDY

We conducted a pilot study to check whether our system could possibly represent higher definition moods and the users could recognize the differences in music. Ten students from IT University of Copenhagen, Denmark volunteered to participate in our pilot study.

As seen in Figure 2, we used Russell's two-dimensional valence/arousal space to locate various types of mood. It is worth noting how these appear in a circular orbit around the origin, this means that the closer we get to the center the more indistinct the mood would result.

We tried to express some of the adjectives through our music generator. After noticing that some are so close in the space that they were very difficult to differentiate (e.g., *astonished* and *aroused*), we finally decided to make eight clips. This decision was also brought forth from the desire of keeping the length of the experiment below ten minutes so that the tester wouldn't get tired and so affect the quality of the data. The final emotions that we selected are bored, content, happy, miserable, tired, fearful, peaceful and alarmed. It should be noted that we have two moods that don't appear in Russell's study: *fearful* and *peaceful*. With these we wanted to express some feelings that are more commonly found in music, as Russel's study was only focused on emotions and not on music.

We defined fearful to be a mood with *medium-low valence* and *medium-high arousal*, which would put it very close to the *frustrated-annoyed-angry* cluster as in Figure 2. Peaceful, on the other hand, has *medium-high valence* and *medium-low arousal*, so it would be part of the *content-satisfied-calm* cluster. With this and Content mood, we could explore if people could see a difference between such closely located moods.

When we defined the moods, we asked the tester to place the mood they feel in the valence/arousal space. We believed this was important as people might have different definitions of the mood adjectives. It should also note that, for the majority of the participants, English was not their first language.

As the valence/arousal space is not something that most people use in ordinary life, we employed the SAM (Self Assessment Manikin) pictures with two sliders (representing a Likert value from one to five), with the texts describing the meaning of the dimensions [13].

In addition to the demographic data of the study participants, such as age and gender, data relating to their music preference (such as genre) and average time for listening to music were gathered in an open way, by having as answers a four point scale: *"always" "often", "seldom", "never"*.

Overall the experiment survey presented: a personal data section, a mood recognition questionnaire (where the participants listened to the clip and then specified which valence/arousal level they felt in the music) and finally a section in which they could select up to two emotion adjectives to describe the piece of music.

As an effort to influence the participants as little as possible with the emotion words, we divided their decision of music recognition in two parts. So while listening to the music, the participants could already set valence and arousal values without seeing the emotion adjectives. Once the clip ended, the last part of the form appeared.

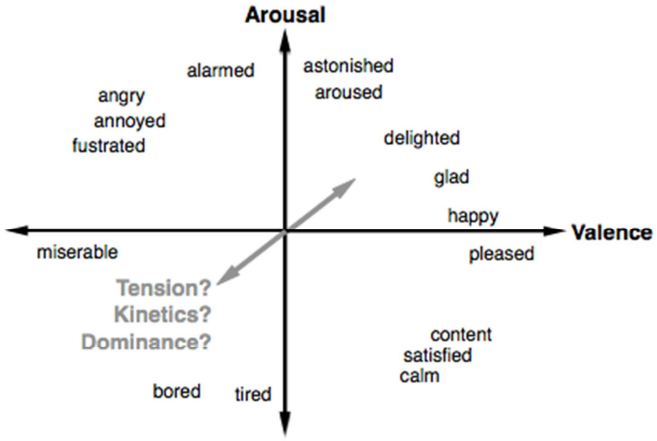


Figure 2: The Valence-Arousal space, labeled by Russell's direct circular projection of adjectives [16]. Includes semantic of projected third affect dimensions: "tension"[5], "kinetics" [15], "dominance" [14]. In our study we haven't considered this third dimension as it's still not very defined.

5. RESULTS

Figure 3 and Table 1 shows the study results including percentage of correct answers by the study participants. The least correct answers were for *happy* music, as low as 30%. The mood types that had highest recognition were **peaceful** and **alarmed** (with an 80% of correct guesses) and **bored** (with 60%).

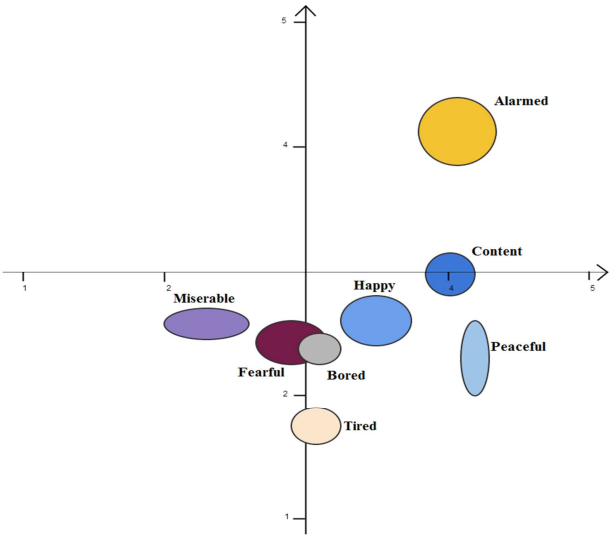


Figure 3: Distribution of the mood types recognized by the participants. (Vertical axis represents Arousal and horizontal axis represents Valence.)

Table 1: Percentage of correct answers by participants

Bored	Content	Happy	Miserable	Tired	Fearful	Peaceful	Alarmed
60%	40%	30%	40%	40%	40%	80%	80%

The mean ratings of arousal and valence for each type of mood music are summarized in Table 2, with a five point scale from 1 (most negative for valence and most calm for arousal) to 5 (most positive for valence and most active/stressful for arousal). For example, we can see how **peaceful** is perceived as a high valence (4.2) and medium/low arousal (2.3) mood, which is very close to what we expected.

Some interesting results are yielded by **fearful** and **alarmed**:

Fearful mood appears to be perceived as a slightly low arousal emotion (2.4), while our expectation was for it to be medium/high.

Alarmed mood is expected to be an almost pure arousal mood with just a very small negative valence (or almost neutral), but it turned out high valence (4.1) with high arousal (4.1). It seems that the participants had no issue in recognizing the arousal component, but found the music to have a positive valence.

Table 2: Average ratings of valence and arousal for various moods (scale from one to five)

	Valence	Valence STD	Arousal	Arousal STD
Peaceful	4.2	0.422	2.3	1.252
Fearful	2.9	0.994	2.4	0.699
Bored	3.1	0.568	2.4	0.516
Content	4	0.667	3	0.667
Alarmed	4.1	1.135	4.1	1.101
Happy	3.5	0.972	2.6	0.843
Miserable	2.3	1.252	2.6	0.516
Tired	3.1	0.738	1.8	0.632

6. CONCLUSIONS

We had some interesting results regarding emotional adjectives. There seems to be a consensus on the semantic meaning of these words is lacking and, moreover, correlations between different emotion words seem to emerge (for example *content*, *peaceful* and *happy*). This made our early analysis seem to have pretty negative results for most moods (apart from *peaceful*, *alarmed* and *bored*). By examining attentively the data, however, we could see how the results were much closer to what we expected. We should also note that the emotional meaning associated to chord sequences differs between individuals (even though some chord sequences have a more shared emotional perception) as found out by Yasuda and Abe [20].

While just early results of our pilot study with a small number of participants seem to indicate some positive results, it also shows us the problem of using emotional adjectives in this particular type of testing. As future work, we plan to conduct our study with more participants, revising the usage of adjectives to express emotions by either eliminating them completely or giving the subjects a definition of them to resolve a possible problem of ambiguousness. Another direction we are looking into currently is the use of digital games to collect ground truth data as games are known to serve as an effective platform to crowdsource data from the user [10][8][2].

7. REFERENCES

Aucouturier, J.-J., Pachet, F. and Sandler M. (2005). "The Way It Sounds": Timbre Models for Analysis and Retrieval of Music Signals

Barrington, L., O'Malley, D., Turnbull, D., & Lanckriet, G. (2009, June). User-centered design of a social game to tag music. In Proceedings of the ACM SIGKDD Workshop on Human Computation *(pp. 7-10). ACM.*

Beedie, C., Terry, P., & Lane, A. (2005). Distinctions between emotion and mood. Cognition & Emotion, 19*(6), 847-878.*

Brown, D. (2012, July). Mezzo: An Adaptive, Real-Time Composition Program for Game Soundtracks. In Eighth Artificial Intelligence and Interactive Digital Entertainment Conference.

Eerola, T., Lartillot, O., & Toiviainen, P. (2009, October). Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models. In ISMIR (pp. 621-626).

Houge, B. (2012) Cell-based music organization in tom clancy's endwar. In Eighth Artificial Intelligence and Interactive Digital Entertainment Conference .

Katayose, H., Imai, M., & Inokuchi, S. (1988, November). Sentiment extraction in music. In Pattern Recognition, 1988., 9th International Conference on *(pp. 1083-1087). IEEE.*

Kim, Y. E., Schmidt, E. M., & Emelle, L. (2008, February). MoodSwings: A Collaborative Game for Music Mood Label Collection. In ISMIR *(Vol. 8, pp. 231-236).*

*Kreutz, G., Ott, U., Teichmann, D., Osawa, P., & Vaitl, D. (2008). Using music to induce emotions: Influences of musical preference and absorption.*Psychology of music, 36*(1), 101-126.*

Law, E. L., Von Ahn, L., Dannenberg, R. B., & Crawford, M. (2007, September). TagATune: A Game for Music and Sound Annotation. In ISMIR *(Vol. 3, p. 2).*

Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). "Expressivity comes from within your soul": A questionnaire study of music students' perspectives on expressivity. Research Studies in Music Education, 20*(1), 23-47.*

Liu, D., Lu, L., & Zhang, H. (2003, October). Automatic mood detection from acoustic music data. In ISMIR.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology, 14*(4), 261-292.*

Mehrabian, A., & Russell, J. A. (1974). An approach to environmental psychology. the MIT Press.

Mion, L., & De Poli, G. (2008). Score-independent audio features for description of music expression. Audio, Speech, and Language Processing, IEEE Transactions on, 16(2), 458-466.

Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39*(6), 1161.*

Scirea, M. (2013). Mood Dependent Music Generator. In Advances in Computer Entertainment *(pp. 626-629). Springer International Publishing.*

Smith, B. D., & Garnett, G. E. (2012, July). Improvising musical structure with hierarchical neural nets. In Proc. AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE).

Thayer, R. E. (1989). The biopsychology of mood and arousal. *Oxford University Press.*

Yasuda T., & Abe J. (2003) Are emotions perceived differently among individuals? Music Perception, 21*(1), 156–157.*

AN INVESTIGATION OF THE CROSS MODAL INTERACTION BETWEEN SONGS AND THEIR ALBUM COVERS IN THE MOBILE DIGITAL ENVIRONMENT

Adam Strange[1], *Atsushi Marui*[1]
[1]*Tokyo University of the Arts, Japan*
Time: 15:00

ABSTRACT

A multi-part subjective evaluation was carried out to determine if album cover artwork has a significant impact on the preference and consumption of music in a mobile-device environment. In the first of two experiments, 66 participants were separated into 3 treatments, one group seeing only the album cover image, one group only hearing the music, and one group receiving both the music and corresponding album cover image. 12 songs from popular, classical, and jazz genres were included. Participants used their own smartphones to display an internet-based GUI designed to simulate the typical conditions of mobile music listening, and reported their level of preference, purchase intention, any previous experience listening to the given song and also to songs similar to the given song. An ANOVA of the results revealed that there was no significant difference between preference ratings between any of the treatments for 11 of the 12 songs. The second experiment confirmed the results of the first experiment using a new set of 12 songs and covers rated by 11 participants who then participated in an elicitation phase where a set of 84 attributes describing the relationship between songs and album covers were recorded. A panel of 4 then revised this list to 29 unique attributes and used them to rate the 12 songs and covers. These ratings were then also submitted to an ANOVA as well as cluster analysis. The results of the second experiment showed no significant difference between the three groups' preference ratings in the first part or the panel's attribute ratings between the songs and covers. Additionally, the cluster analysis suggests that the original 29 rating scales could be reduced to a representative set of 8. It is concluded that album cover artwork is overall highly representative of the music contained within on both an attribute and a preference level, but that it does not influence impressions of songs in a significant way. Also, it is suggested that ratings on an attribute from each of the 8 clusters could be used to gauge the goodness-of-fit for new songs and would-be album covers prior to dissemination on digital marketplaces.

1. INTRODUCTION

In recent years, the amount of music listened to and purchased on mobile devices has greatly increased [2]. This environment and medium is considerably different from previous music consumption methods. One such difference is the loss of physical media. However, the image of the physical media's front is generally retained in the digital distribution. This retention of cover image raises the question of whether the image is a necessary element in the appreciation of any given song, and thus is indispensable, or whether its purposes are purely commercial. In the first of two experiments, musical preferences according to the presence or absence of an album cover image were evaluated. In the second experiment, the relationship found in the first experiment regarding overall preference was further investigated by eliciting attributes to uncover a more structural relationship between songs and their album covers.

2. EXPERIMENT 1

Since the goal of this research is to investigate the appreciation of music specifically in the context of mobile-based digital distribution, it was decided that actual smartphones should be utilized to properly simulate actual-use conditions. By creating an internet-based GUI, it was possible to allow the participants to use their own personal smartphones, to which they have grown accustomed to and feel comfortable with, while reducing any hassle or problems with app installation that would otherwise be necessary. Using the Internet also allowed the participants to participate in the experiment at a time and place of their choosing. While this is contrary to experimental designs that control the surroundings in which the experiment is executed, it provides the benefit of increased ecological validity [5]. The participants were told in advance that they would be listening to music and rating their preferences, and it is assumed that they did so in the very same environments that they would ordinarily access music on their mobile devices. However, to combat the increased variance brought by disparate locations, it was necessary to ensure that recruitment numbers were sufficiently high. This is particularly the case for the first experiment.

2.1. Stimuli

12 songs and their corresponding album cover images as available through digital distribution were selected. The songs were chosen at random from 4 different lists, Billboard's U.S. Hot 100 number one singles of 2012 [8], Billboard's Japan Hot 100 singles number one tracks of 2012 [7], National Public Radio's Top 10 Jazz Recordings of 2012 [10], and National Public Radio's Top 10 Classical Recordings of 2012 [9]. The two Billboard lists provide music that achieved high levels of commercial success, while the two NPR lists provide variety via critically acclaimed tracks of two genres that do not generally find their way to the Hot 100. A wide variety of musical colors and styles are spread across the 12 tracks, which are listed in Table 1. In addition to the 12 main tracks, an additional song and its album cover were selected from each list to be used as dummy tracks to ensure all experimental treatments had music and images to rate (See Treatments below). Each song was presented as a 90 second preview in the style of the iTunes MP3 store, and loudness across tracks was equalized according to Apple's SoundCheck algorithm. While this is not considered to be the most accurate algorithm available, it may be more appropriate for mobile devices [3]. It is highly likely to be employed in actual-use scenarios at the consumer level, and an informal check of the stimuli in use showed even results. The stimuli were presented in random order, with the rule that the dummy tracks be the last 4 presented.

2.2. Rating Scales

The participants rated the stimuli across 4 rating scales. The question and answer formats of each are shown in Table 2. Since the participants would not be receiving training, and the vast majority would not have experience in sensory attribute rating, it was decided that overall preference would be the most appropriate rating scale, in accordance with the discussion provided by Barrios and Costell 2004 [1]. Also, a rating scale polling purchase intent was included to insure ecological validity in the context of digital music consumption [5]. Additionally, two rating scales regarding listening experience, one for the given song and one for any experience with songs similar to the given song, were included on the basis that they showed potential to influence preference according to the authors' previous research [17].

2.3. Treatments

The participants were split into one of three treatments, each treatment receiving a different variation of the same stimuli. The first treatment was presented with both a song and the song's album cover image simultaneously. The second treatment was presented only the song, and the third treatment was presented only the image. As such, each participant was presented with only one variation of the stimuli. This strategy was adopted from research on cross modal interactions carried out at McGill University in 2011 by Vines, et al [15]. As a result, the data analysis must focus on differences across participants rather than within participants, but effectively removes the potential order bias that would occur from the same participants being presented with multiple versions of the same stimuli.

2.4. Participants

Participants were recruited in three ways, graduate and undergraduate students in the Musical Creativity and the Environment program at Tokyo University of the Arts, respondents to invitations to participate via SNS Facebook, and respondents to flyers posted at a brick-and-mortar store. A total of 66 participants successfully completed the survey, resulting in 21 responses for the song & image treatment, 20 responses for the song-only treatment, and 25 responses for the image-only treatment.

2.5. Procedure

The participants accessed the survey online using their personal smartphones and head/earphones via a URL supplied to them after responding to the recruitment invitation. The participants were allowed to access the survey at any place and time of their choosing with the direction that they would be listening to music (song&image or song-only treatments) or viewing album covers (image-only treatment) and rating how much they like them. The participants were not aware that there were other treatments with different versions of the stimuli. The survey began with an explanation/instruction phase, which included headphone volume calibration before beginning their responses. The four dummy tracks were added to the end of the experiment for all treatments, and were presented simultaneously with their album covers. This was done primarily to ensure that the image-only group took the survey with headphones on and was in a music-listening mind-frame to bring parity with the other two treatments.

2.6. Results

A two-way, mixed-model ANOVA was carried out on the preference data, with treatment as a between subject factor and song as a within subject factor. The resulting *p*-values show statistical significance for both treatment (0.025) and song (0.0001), and additionally the interaction between them (0.017). At first glance, it appears that the album cover image is indeed important; however, a more in-depth post-hoc is necessary to reveal the full story.

Using Ryan's method for the ANOVA post-hoc, the significance of each treatment and song can be seen individually. As can be seen in the top section of Table 3, only one of the twelve songs achieves statistical significance. Regarding the significance between the three treatments, it can be seen in the bottom half of Table 3 that the statistical significance for treatment appears exclusively between the song&image treatment and the image-only treatment. Thus, besides the ratings between those two treatments for the 6th song only (*Young, Wild & Free*), the differences in preferences do not reach significance even in comparison of the song-only ratings vs. the image-only ratings.

Regarding the ratings of purchase intent, Spearman's rank correlation coefficient was calculated to asses the relationship between purchase intent and preference with a result of 0.76 demonstrating a clear correlation. The two ratings on experience were also explored in the same way, but yielded low results of 0.13 for experience with the given stimuli and 0.07 for experience with songs/images similar to the stimuli, thus demonstrating a clear lack of correlation between experience and preference.

2.7. Discussion

Surprisingly, the results indicate that preference ratings for any given song and its cover are much the same overall. Due to the experimental design incorporating different subjects in the three treatments, it is not possible to rule out disproportionate musical tastes amongst the groups as a possible factor from this data alone. Also, it is difficult to label *Young, Wild & Free* an outlier from only 12 tracks worth of data. However, there is clear indication that songs and their covers are not arbitrarily matched, and that they could in fact share a variety of intricate information thus leading to the preference link. Considering the experience ratings lack of correlation, it would appear that covers can relay data to potential listeners even if that listener is not highly familiar with the genre at hand. This suggests that album covers portray some kind of abstract and highly universal information.

To shed light on the above findings, a second, multi-step experiment was conducted using a within-subjects design for the treatments and adding the elicitation of attributes then used as rating scales to explore the connection between songs and their covers in depth.

3. EXPERIMENT 2

The following experiment was carried out in 2 main parts with 3 primary objectives. The first objective was to confirm the results of the previous experiment related to the comparison of album cover and music preferences. The second objective was to elicit a set of descriptors to be used to evaluate the characteristics shared between songs and album covers. The final objective was to use those descriptors as a rating device through which to evaluate the constructs shared between songs and their album covers. This was done by constructing a 4-person panel to handle the attribute refinement and use the attributes to rate the stimuli.

3.1. Stimuli

12 songs and their corresponding album covers were chosen in a very similar manner to the previous experiment. Where as the previous experiment used 4 lists from 2012, this time the corresponding lists from 2013 were used. The lists were: Billboard's U.S. year-end Hot 100 singles [11], Billboard's Japan Hot 100 number 1 singles [12], NPR's top 10 Jazz tracks [13], and NPR's top 10 classical tracks [14]. 3 songs were selected at random from each list along with the corresponding album art image as included in the commercially available digital distribution of each track. Additionally, 3 album cover images were randomly chosen from the first experiment's stimuli. Thus, the number of images outnumbered the number of songs 15 to 12. This was done in an attempt to prevent participants from trying to match songs to album covers which would have resulted in a strong order-bias. The data for the extra 3 images were not included in the analysis.

The songs were loudness equalized according to Apple's SoundCheck algorithm. Song playback began at the beginning of the track and was truncated to 2 minutes in length. This allows a unified starting position amongst all the songs, and is more representative of the playback conditions of services such as YouTube [15] or Sony's Music Unlimited [16], where tracks or videos begin at the beginning without a sound preview available. The 2-minute duration ensured conservative bandwidth usage, which was deemed important since the participants would be using their own personal smartphones to access the GUI.

For the initial preference rating of Part 1, songs and album covers were presented separately in random order for a total of 24 stimuli. For the elicitation phase of Part 1, each song was presented together with its corresponding album cover image in random order for a total of 12 stimuli.

3.2. Rating Scales/Input

Part 1

The only principle rating scale answered by participants in Part 1 was preference, phrased as "How much do you like this song?" or "How much do you like this album cover?". In order maximize the effectiveness of ANOVA as an analysis tool, the previous point-based rating scale system was replaced by a continuous rating scale. The continuous rating scale allows the data to better meet ANOVA's necessary conditions [4].

The scale was incorporated into the GUI as a typical slider interface, as often used for the setting of brightness and volume controls on smartphones. To ensure that the participants felt comfortable using the sliders, a brief explanation and training phase was added to the beginning of the test. All participants generated highly favorable results in the slider training phase, demonstrating their ability to effectively use the slider as a rating system. For the preference evaluation, the left pole was labeled "Highly Like", and the right pole was labeled "Highly Dislike". The slider position defaulted to the middle.

The elicitation phase was loosely inspired by pair-wise applications of the repertory grid technique. Pairs of stimuli were compared or contrasted by the participant via a descriptive word or phrase in an adjectival format. However, stimuli pairs were limited to a song paired with its corresponding album cover image. A question phrased, "Does this song match this album cover?" was displayed under the stimuli pair with multiple-choice answers "Yes" or "No". Based on the participant's answer, a second question was displayed below it, either "Why do they match?" or "Why don't they match?". For matching pairs, participants filled in the blank to the sentence "They both are…", or for mismatched stimuli, "The song is…, but the cover is…". In this way, one attribute was elicited for each stimulus pair thought to be matching, and a pair of antonyms was elicited for each stimulus pair thought to be mismatched.

Part 2

84 unique words were elicited during part one. In Part 2, a panel then reviewed the output and discussed the uniqueness of the original 84 attributes, ultimately reducing the list to 29 attribute scales. The attributes were then rated in intensity on a simple 1 to 5 scale for the 12 songs and covers.

3.3. Subjects

11 undergraduate and graduate students in the Musical Creativity and the Environment program at Tokyo University of the Arts participated in Part 1, 6 female and 5 male.

In Part 2, the 4-person panel was recruited from graduate students in the same program, two of whom participated in Part 1, 1 female and 3 male.

3.4. Procedure

In Part 1, participants accessed the test GUI via their personal smartphones and headphones at any time and location of their choosing as in the previous experiment.

Participants were asked to rate their preferences for song-only and cover-only stimuli via smartphone in the same way as the previous experiment. The participants were explicitly informed that there were 15 images and 12 songs to be rated in an attempt to prevent them from trying to match song-stimuli with image-stimuli. In informal talks with the participants after the survey, they were not aware that all the song-stimuli did in fact have a corresponding image during the first section. Unlike the first experiment, a song&image treatment was not presented.

Subjects were then presented with a song paired with its according album cover image and asked whether or not the song and album cover matched. If they felt the two matched, they were asked to write in an adjective describing why they matched. Conversely, if the subject felt that the song and cover did not match, they were asked to write-in two miss-matched adjectives, one describing the song, and one describing the cover.

In the second half of the experiment, a panel of four participants reviewed the previously elicited attributes. The panel then reduced the list to a set of attributes with unique meanings. A series of two 2-hour sessions to refine the attribute list and then discuss and come to a consensus on the meaning and use of each of the 29 attribute scales for both images and songs were carried out. The panel then used the attribute set to evaluate the songs and album covers from the previous half.

3.5. Results

The results of the preference evaluation in Part 1 were analyzed via ANOVA to confirm or disconfirm the findings of the first experiment. A 2-way ANOVA was used with the within-subject factor of treatment (Audio or Image) and the within-subject factor of song. The treatment factor did not produce statistical significance with a p value of 0.17. The song factor did reach statistical significance with a p value of < 0.0001. The interaction between song and treatment was not statistically significant with a p value of 0.95.

The data from the 29 attribute scales from Part 2 were also analyzed by an ANOVA. This time, treatment and song were within-subject variables and attribute was added as a blocking variable. Treatment and its interaction with song both did not yield statistical significance ($p = 0.13$, $p = 0.78$). Song and attribute did reach statistical significance with p values of 0.009 each.

In order to get an image of the use of the rating scales and to further refine them for future studies, the ratings, averaged across the four panelists, were submitted to cluster analysis using Euclidean distance and Ward's method. The resulting dendrogram is shown in Figure 1, with the red outlines representing an 8-cluster solution. If the ratings for a given attribute are similar between the songs and images, then the two versions of the rating scales will appear in the same cluster or at least appear in adjacent clusters. In this case, the dendrogram shows good agreement between the song and image versions of the scales, with 25 of the 29 pairs found in the same or adjacent clusters. This gives us a quick visual confirmation of the results obtained via ANOVA, and also suggests that there are 8 main areas in which songs and their album covers share abstract information.

3.6. Conclusion

The results of the Part 1 ANOVA are in line with the results of the first experiment and confirm that there is no significant difference between the preference ratings between songs alone and their cover images alone. Furthermore, the ANOVA from Part 2 demonstrates that there is also no significant difference between the attribute ratings of the songs alone and cover images alone. This reinforces the suggestion that songs and

their album covers share attributes at an abstract level, and that album covers are able to inform potential listeners about the musical contents in a way that allows them to estimate preference level and make assumptions about the affect.

Using the 8 clusters derived from the attribute ratings, a variety of potential applications are opened up. Perhaps the most straightforward would be to use the rating scales to access the appropriateness of a set of songs and their potential album cover before release to ensure a good fit. Furthermore, listeners could potentially receive suggestions of music they may like based on their preference ratings of album covers alone. Additionally, listener preference ratings could be combined with panel ratings via an analysis method such as preference mapping [1].

The scope of this particular research is bound to album covers, but with further research, it may be possible to increase the boundaries of the attribute scales and preferences to include more than just album covers. This would pave the way for a better understanding of the cross-modal interplay between musical and visual information overall, and could make a useful contribution in a variety of different fields.

4. REFERENCES

[1] Barrios E.X. and Costell E. "Use of Methods of Research into Consumers' Opinions and Attitudes in Food Research." In Food Science and Technology International, 10(2004): 359-371
[2] David Kusek and Gerd Leonhard. The Future of Music: Manifesto for the Digital Music Revolution. Boston: Berklee Press, 2005
[3] Lund, Thomas. ITU-R BS.1770 Revisited. Accessed 2013/04/23, http://www.tcelectronic.com/media/1017421/lundt013011.pdf
[4] Miller, Rupert G. Jr. Beyond ANOVA: Basics of Applied Statistics. New York: Wiley, 1986
[5] Schmukler, Mark A. "What is Ecological Validity? A Dimensional Analysis." In INFANCY, 2/4(2001): 419–436

[6] Vines BW, Krumhansl CL, Wanderley MM, Dalca IM, Levitin DJ. "Music to my eyes: cross-modal interactions in the perception of emotions in musical performance." In Cognition, 118/2(2011):157-70
[7] List of Hot 100 number-one singles of 2012 (Japan). Accessed 2013/04/09, http://en.wikipedia.org/wiki/List_of_Hot_100_number-one_singles_of_2012_(Japan)
[8] Billboard Year-End Hot 100 singles of 2012. Accessed 2013/04/09, http://en.wikipedia.org/wiki/Billboard_Year-End_Hot_100_singles_of_2012
[9] Top 10 Classical Albums of 2012. Accessed 2013/04/13, http://www.npr.org/blogs/bestmusic2012/2012/12/04/166505256/ top-10-classical-albums-of-2012
[10] Top 10 Jazz Albums of 2012. Accessed 2013/04/13, http://www.npr.org/blogs/bestmusic2012/2012/12/07/166760734/top-10-jazz-albums-of-2012
[11] Billboard Year-End Hot 100 singles of 2013. Accessed 2014/03/15, http://en.wikipedia.org/wiki/Billboard_Year-End_Hot_100_singles_of_2013
[12] List of Hot 100 number-one singles of 2013 (Japan). Accessed 2014/03/15, http://en.wikipedia.org/wiki/List_of_Hot_100_number-one_singles_of_2013_(Japan)
[13] The 2013 NPR Music Jazz Critics Poll. Accessed 2014/03/15, http://www.npr.org/blogs/bestmusic2013/2013/12/16/251761858/the-2013-npr-music-jazz-critics-poll
[14] NPR Classical's 10 Favorite Albums Of 2013. Accessed 2014/03/15, http://www.npr.org/blogs/bestmusic2013/2013/12/13/249751178/npr-classicals-10-favorite-albums-of-2013
[15] YouTube. Accessed 2014/06/06, http://www.youtube.com
[16] Music Unlimited. Accessed 2014/06/06. https://music.sonyentertainmentnetwork.com
[17] Strange, Adam. An Investigation Into The First Impressions Of Jazz Music: The Subjective Evaluation Of The Modern Listener, Tokyo University of the Arts, Music Research Division Master's Thesis 2012

SALIENCY-DRIVEN MODEL FOR PERCEPTUAL AUDIO ONSET DETECTION

Sungkyun Chang[1], Kyogu Lee[2]
[1]Music and Audio Research Group, Graduate School of Convergence Science and Technology, Seoul National University, Korea, Republic of,
[2]Music and Audio Research Group, Graduate School of Convergence Science and Technology, Seoul National University, Korea, Republic of
Time: 15:30

1. Background and Aims: In music signal processing, previous onset detection study has been mainly focused on finding precise time-locations of ALL audio events. Reminding that our brain achieves audio event with selective attention, the conventional onset detectors may fail to match with human onset response in specific cases. This paper describes a model for perceptual onset detection.

2. Method: In experiment, we compare the proposed model with conventional one using synthetic dataset. The dataset consists of synthesized musical notes and recorded environmental sound. We generate it in a similar way to Kaiser's auditory attention examples, such as temporal masking, gap-in-noise.

3. Result and Conclusions: The result showed that the proposed model minimized errors for all the cases. Thus saliency-driven model performs better than conventional one.

[11C] Sound beyond Media

MONTREAL 14:00-16:00 Friday 8 Aug 2014

SOUND QUALITY ENHANCES THE MUSIC LISTENING EXPERIENCE

Victoria Williamson[1], Michael South[2], Daniel Müllensiefen[2]
[1]University of Sheffield, United Kingdom, [2]Goldsmiths, University of London, United Kingdom
Time: 14:00

ABSTRACT

Most people are unable to discriminate the sound quality of music recordings across a range of typical commercial products. However, the effects of very high digital resolution levels on the music listener remain untested. The present study aimed to determine whether people respond differently to recorded music at three resolution levels: 320kbps (mp3), 16-bit/44.1kHz (CD), and 24-bit/192kHz (Studio Master).

In Study 1, participants rated the sound quality of continuous 30-second song segments, presented randomly across the three resolution levels. We replicated the well-established failure to discriminate mp3 and CD, but noted a preference for Studio Master. In Study 2, participants scored the sound quality of two songs at the lowest (mp3) and highest (Studio Master) digital resolution levels. We also assessed felt emotion and analyzed listening time as an implicit measure of preference. In all cases participants successfully discriminated between the two resolution levels and preferred Studio Master recording. Implications for future tests of sound quality in music recordings and music listening practices are discussed.

1. INTRODUCTION

The history of music recording has been marked by the evolution from analogue to digital formats. Originally, music recordings were made by processes that transformed air molecule vibrations into analogous ('analogue') physical alterations in the reception medium, such as vinyl LP grooves. Modern digital recordings, by comparison, convert the pattern of source vibrations into sequences of 1s and 0s. Modern digital sound recording formats incorporate 'lossy' psycho-acoustic algorithms that allow music file size to be compressed, thereby accommodating portable, low cost music listening technology.

Digital recording of music and subsequent compression is a complex process. Two parameters that determine the resolution of auditory information and that are relevant to the present paper are sampling rate and bit depth. Sampling rate is the number of times per second that the original continuous signal is sampled. For example, CD music is encoded at 44.1kHz, meaning that 44,100 samples per second are extracted from the original signal. Bit-depth determines the range of different values possible when measuring the amplitude of the analogue waveform. For example, the 16-bit depth of CD recordings offers 65,536 discrete values of amplitude. Together, the sampling rate and bit depth describe the digital resolution or 'sound quality' of recording, which for CD is 705.6 kilobits per second per channel (kbps; sampling rate x bit depth).

In today's music market the most consumption is via highly compressed digital formats, namely mp3 (Krause, North & Hewitt, 2013). The highest sound quality level supported by mp3 is 320kbps, which equates to a compression ratio of around 4:1.

The benefits to the listener of flexible access to thousands of pieces of music in a miniature device are beyond doubt. However, assuming music is played on devices that are capable of reproducing high digital resolution the question remains; does higher sound quality (digital resolution) lead to a richer experience for the music listener?

Research studies have taken one main approach to tackling this question: testing whether people can discriminate between mp3 and CD levels of resolution. Results in this area are consistent and suggest that people are insensitive to the difference. Pras, Zimmerman, Levitin and Guastavino (2009) evaluated CD and mp3 discrimination using pair-wise comparison of 5 -11 second samples of pop, rock, orchestral, contemporary and opera music. Thirteen 'expert' participants (musicians, studio engineers) could not discriminate between CD and mp3 (256kbps and 320kbps levels). Similar negative results were obtained by Pras and Guastavino (2010) and Yoshikawa et al. (1995), although it should be noted that Olive (2011) found that 70% of their student participants and 86% of 'expert' listeners preferred CD to the most highly compressed mp3 (128kbs) sound.

In most cases it seems that standard CD and mp3 formats are difficult to discriminate, even for even highly trained listeners. However, there is suggestive evidence from Olive (2011) that larger differences in digital resolutions yield above chance levels of detection and a preference for higher resolution.

Developments in digital audio technology mean that it is now necessary to extend the standard comparisons of sound quality in music recording to include Studio Master along with traditional mp3 and CD resolution. Studio Master music recordings are typically encoded with a bit-depth of 24-bit and very high sample rates, such as 192kHz (i.e. 4608 kbps), as is the case of the stimuli of the present study.

Psycho-acoustic studies provide evidence that people may be able to discriminate between CD (705.6 kbps) and Studio Master (4608 kbps) digital resolution thanks to the higher sampling rate. Woszczyk (2003) argued that the temporal resolution for CD (sampling interval of 22.7μs) is insufficient, given that many transient onsets possess a rise time of less than 10μs. By comparison, the 5.2μs sampling interval of Studio Master recording resolves more waveform detail. The human ear could thus discriminate between these two resolution levels if it were able to resolve intervals of c.10μs. Krumbholz, Patterson, Nobbe and Fastl (2003) used backward masking to identify temporal resolution thresholds in the human ear of between 10μs and 20μs.

More recent findings point to another possible mechanism of higher sound quality discrimination: the presence of inaudible high frequencies. Kuribayashi, Yamamoto, and Nittono (2014) tested double blinded discrimination of two short musical excerpts (200 seconds of J. S. Bach's French Suite No. 5) with or without inaudible high-frequency components (above 20 kHz). Participants were unable to identify the difference in these two short samples however, high-α EEG power (10.5-13 Hz) was larger for the last 50 seconds of the excerpt with the high-frequency components than for the excerpt without them. Therefore, inaudible high frequencies, of the kind present in higher sound quality recordings, have an impact on brain activity in the absence of conscious awareness.

These combined findings suggest that the higher sampling rates of Studio Master recording levels might result in differences in physiological response to and/or perceptual experience of music compared to lower levels (mp3 and CD) tested to date.

1.1. The present study

The present study built on previous evidence and investigated discrimination of music at three digital resolution levels: mp3 (320kbps), CD (705.6 kbps) and Studio Master (4608 kbps).

Firstly, we expected to replicate poor listener discrimination between mp3 and CD levels, but we hypothesized that the highest sound quality resolution level (Studio Master) would be detectable compared to the lowest level (mp3).

Secondly, the present paper went beyond measuring sound discrimination since the 'listening experience' is not limited to this dimension. The importance of emotional reactions in music listening is paramount since '*The most common goal of musical experiences is to release emotions*' (Juslin & Vastfjall, 2008, p.559). To date however, we found no research into how sound quality impacts on the emotions felt by the

music listener. In order to measure felt emotion in the present study we employed the Geneva Emotional Music Scale (GEMS-9) (Zentner et al., 2008).

Finally, it is important to note that a music listener's experience extends beyond conscious evaluations related to either discrimination or felt emotion. Implicit responses have been recorded successfully in consumer behavior studies, where participants react to music without being consciously aware of its concomitant properties or, in some cases, even its presence (North & Hargreaves, 2008).

The implicit measure of the listening experience considered in the present study was self-determined listening time, a proxy measure of preference. Lindsen, Moonga, Shimojo and Bhattacharya (2011) investigated sampling behaviours and decision-making ("like" or "dislike") for closely matched musical excerpts. In this study people were free to sample excerpts until reaching a decision on their preference. Participants listened to their preferred option ("like") for longer when making their decision in 73% of trials in contrast to a limited sample bias for the "dislike" condition. People were unaware that their listening time was being measured, thereby providing an implicit objective measure of listening preference.

In summary, the present paper compared the impact of three sound quality levels (mp3, CD and Studio Master resolution) on the 'listening experience' as defined by explicit discrimination, levels of felt emotion, and time spent listening. Study 1 comprised explicit rating of continuously changing sound quality in a single song across all three levels (mp3, CD and Studio Master). Study 2 comprised explicit rating of the sound quality of songs reproduced at the two extreme digital resolution levels (mp3 and Studio Master) combined with judgments of felt emotion and an implicit measure of listening preference (listening time).

2. METHOD

2.1. Materials

Both studies in the present paper took place in a sound attenuating booth. Inner and outer chambers included a 102mm thick acoustic modular panel, separated by an air gap of 100mm. Tested in accordance to ASTM standards, the booth provided minimum noise reduction of 37db and a sound absorption coefficient of 0.38 for the 63Hz octave band, rising to 93db noise reduction and a sound absorption coefficient of 1.06 for the 1kHz octave band, and >93db noise reduction and >1 sound absorption coefficient for >1kHz octave bands.

The music reproduction equipment consisted of a Toshiba laptop providing the digital stream, which was fed via an Ethernet router to a Linn Majik DSM streamer (integrated amplifier), connected via Linn interconnects to Linn Majik 109 loudspeakers. Volume was kept at a constant level, assessed in pilot testing to be a comfortable average listening volume for someone of normal hearing. The system employed the Linn Kinsky/UPnP control systems allowing interface to a data logger for recording of listening time (Study 2).

Four music recordings were used across the two studies. All were sung by Carol Kidd accompanied by guitarist Nigel Clark ("Stormy Weather", "I Got Lost in His Arms", "The Shadow of Your Smile" and "Moon River"). All songs were originally recorded at Studio Master level. The recordings were then prepared to different levels of digital resolution using Magix Sequoia software. To create the CD and mp3 files, the original Studio Master recordings were down-sampled to the appropriate standard sample rates and bit-depths.

For Study 1, "Moon River" was prepared as a looped version (song repeated once, duration of 7 min 48s) consisting of 30-second segments. Each 30s segment was followed by an unnoticeable shift in resolution to the subsequent segment with a different and randomly selected digital resolution; all possible combinations of changes in resolution levels were included. The final 'track' was rendered and exported at the maximum quality of 192kHz/24-bit. The lower quality samples within the track (CD and mp3) were still of low quality after the rendering treatment and high quality export, since digital information had been lost in the down-sampling. The resultant single 'track' presented a seamless repeated song to the listener.

In a break between the two studies participants heard "The Shadow of Your Smile" (Duration: 4 min 49s) in mp3 format. Like the other habituation phases of the procedure the use of mp3 music in this way helped to establish a constant baseline level of recent music exposure for all participants. For Study 2, "Stormy Weather" (Duration: 3 min 47s) and "I Got Lost in His Arms" (Duration: 3 min 37s) were prepared at mp3 and Studio Master levels using the above techniques.

The GEMS-9 was used to assess felt emotion in Study 2. The GEMS-9 presents nine factors with emotion adjectives. These include Wonder (Dazzled, Moved), Tenderness (Affectionate, In love), and Nostalgia (Dreamy, Melancholic). Participants rate each of the nine factors on a 5-point scale from felt "Not at all" to felt "Very much" (Zentner et al., 2008).

2.2. Participants

In total 20 participants (8 female) took part in the studies. Participants were graduate students aged between 21 and 59 with a mean age of 27.8 years (SD = 8.5). No reward was offered for participation.

2.3. Procedure

Participants were seated centrally in the booth, equidistant from each loudspeaker. We began with a habituation phase where participants completed a musical experiences questionnaire (not reported in this paper) while the mp3 version of "Moon River" played in the background. Participants were instructed to pay no attention to the music but to concentrate upon the questionnaire. The aim of the habituation phase was to establish a baseline for the lowest level of sound quality, in case participants had been listening to music recorded at different levels before the study began.

Participants were then given the Study 1 score sheet, and instructed to listen carefully to the repeated version of "Moon River". They were informed that each version of the song, although seamless, consisted of seven 30s segments at differing 'sound quality' (three mp3, two CD and two Studio Master). They were requested to rate each 30s segment for sound quality on a 1-7 scale. Immediately at the point of each 30s change, visible on the playback device, the participant scored the previous segment for sound quality. They were advised not to score the first segment at 1 or 7 in order to avoid floor/ceiling effects, and to score based on their idea of sound quality, not on performance or enjoyment at any particular point. We pointed out that differences in sound quality would be subtle but the importance of making a decision as soon as possible.

Study 2 began following two minutes of relaxation. First participants completed another habituation phase during which they filled in a personality questionnaires (not reported in this paper) while "The Shadow of Your Smile" played at mp3 level in the background. As in Study 1,

the habituation phase served to establish an equal baseline of low sound quality for all participants. Participants were then given the Study 2 sound quality score sheet and told that they would hear two different songs. Their first task was to score each whole song for level of sound quality on the same 1-7 scale as in Study 1. Each participant was played one of the four order-counterbalanced combinations of the two songs ("Stormy Weather" and "I Got Lost in His Arms") at both levels of digital resolution (mp3 and Studio Master).

Participants were informed that they would listen to each song once. After that they were given a remote control and could replay/fast forward/rewind the song as much as they liked in order to make their decision on sound quality. In addition, participants were asked to consider the emotional effect of each song and completed the GEMS-9 for felt emotions. They were asked to rate only the emotions they felt while listening to the music, not emotions conveyed by the music. Again, they were told that they could listen as long as they liked in order to complete the GEMS-9. Total listening time for each song was recorded covertly on the laptop.

3. RESULTS

3.1. Study 1

For the song Moon River (played twice), each participant provided four ratings for CD conditions, four for Studio Master, and six for mp3, generating a dataset of 280 total observations. The means and standard deviations for the three conditions are given in Table 1.

In order to determine whether the raw scores represented in Table 1 were statistically significant, we applied linear mixed models with random effects that account for heterogeneous variances across the groups (R package 'nlme').

Table 1. Means and standard deviations for sound quality ratings across the three levels of digital resolution. Ratings were given on a 1-7 scale, with 7 indicating the highest quality level.

Sound Quality Level	Mean Rating	SD
mp3	4.30	1.03
CD	4.24	1.09
Studio Master	4.63	1.13

For fixed effects we included as independent variables; i) the track position of the sound quality condition, as compared to other segments of the same quality ('Position'), ii) whether it was the first or the second play of the track ('Play'), and iii) sound quality level of the 30s segment (mp3, CD, Studio Master). Participants were included as a random effect and variances for the three sound quality conditions were estimated separately to account for the difference in sample size and variance. Subjective ratings of sound quality were the dependent variable.

The results can be seen in Table 2. The difference between the Studio Master and CD conditions was significant ($t(255) = 2.67$, $p = .008$). When mp3 was used as the reference level, there was a significant difference between Studio Master and mp3 ($t(255) = 2.32$, $p = .021$: not shown in Table 2). None of the remaining effects reached the common significance level.

Table 2. Model parameters and significance levels for mixed effects model of explicit sound quality ratings. In this analysis the model reference levels were CD (Sound Quality), First Position, and the first play of each track

Variable	Parameter Estimate	95% Confidence Intervals		t-value	p-value
		Lower	Upper		
Sound Quality					
mp3	0.05	-0.23	0.33	0.35	.729
Studio Master	0.40	0.10	0.70	2.67	.008
Position					
Two	-0.05	-0.29	0.19	-0.40	.689
Three	0.01	-0.36	0.38	0.07	.944
Play					
Second	0.19	-0.03	0.41	1.72	.087

3.2. Study 2

The results of Study 1 indicated that explicit ratings of sound quality were significantly higher for Studio Master compared to CD and mp3, while the ratings of the latter two did not differ from each other. The first aim of Study 2 was to replicate this main effect, using different experimental stimuli (different songs) and presentation paradigm (whole songs rather than 30s excerpts).

The second aim was to assess whether participants' felt emotion ratings and self-determined listening time would differ between sound quality conditions. The means and standard deviations for all three dependent variables are given in Table 3.

Table 3. Means (standard deviations) for explicit sound quality ratings, listening time, and GEMS score across the two levels of sound quality.

Sound Quality Level	Explicit Rating	Listening Time	GEMS Score
mp3	4.35 (1.14)	179.65 (57.53)	18.2 (4.07)
Studio Master	5.60 (0.99)	211.40 (72.71)	20.7 (4.78)

According to the raw mean figures, participants rated sound quality as higher, felt more emotions, and listened longer to tracks in the Studio Master compared to the mp3 condition. These differences were subsequently tested for statistical significance. To analyze sound quality and emotion ratings we employed two linear mixed-effect models. The two models used sound quality as the main independent variable but also accounted for fixed effects from presentation slot (first vs. second) and song ("Stormy Weather" vs "I Got Lost") while including participants as a random factor.

As can be seen in Table 4, sound quality ratings for Studio Master were significantly higher than those given to mp3 ($p = .001$), while there was no significant difference between the tracks presented first or second nor between the two songs (though this effect bordered on significance;

p = .054). The mixed model for emotions (Table 5) found an advantage for Studio Master compared to mp3 (p = .008), this being the only significant effect in the model.

Table 4. Model parameters and significance levels for the two mixed effects models, for explicit sound quality ratings. In this analysis the reference levels were mp3 (Sound Quality), First Play, and the song "I Got Lost".

Variable	Parameter Estimate	95% Confidence Intervals		t-value	p-value
		Lower	Upper		
Sound Quality					
Studio Master	1.17	0.52	1.82	3.81	.001
Play					
Second	-0.13	-0.79	0.51	-0.43	.670
Song					
Stormy Weather	-0.63	-1.27	0.01	-2.06	.054

Table 5. Model parameters and significance levels for the two mixed effects models, for GEMS emotion scores. Reference levels match those of Table 4.

Variable	Parameter Estimate	95% Confidence Intervals		t-value	p-value
		Lower	Upper		
Sound Quality					
Studio Master	2.67	0.79	4.55	3.00	.008
Play					
Second	-0.68	-2.55	1.19	-0.77	.452
Song					
Stormy Weather	-1.08	-2.95	0.79	-1.22	.239

There was an issue with the analysis of listening time. Many participants opted to listen only up to the end of the song despite the instruction to listen for as long as they liked beyond then. Therefore, listening times were skewed towards 217 and 227 seconds (the play times for the two songs). Because of this skewed data distribution it was not possible to compute a linear model for listening time. Instead, we recoded listening times into three categories: early termination / end of song / beyond end of song and computed a chi square test between sound quality and listening time, using participants as a blocking factor. The association between categorized listening time and sound quality was not significant (chi square = 1.897, p = 0.138).

4. DISCUSSION

For many years it has been believed that very few people can tell the difference between the standard commercially available sound resolution levels, namely CD and mp3. The results from Study 1 support this assertion: sound quality ratings from CD and mp3 resolutions did not differ, in line with previous literature (Yoshikawa et al., 1995; Pras et al., 2009; Pras & Guostavino, 2010).

However, the development of music reproduction technology and a trend towards commercial availability of high resolution recordings, such as Studio Master, has stimulated new research questions – can people detect a difference in sound quality between lower (mp3) and much higher digital resolution (Studio Master) music recordings? Moreover, does such an increase in resolution impact on the music listening experience in terms of felt emotions and/or listening preferences.

Participants in the present study consistently rated Studio Master music as higher in subjective sound quality compared to mp3, both in terms of 30s excerpts in a continuous song (Study 1) and across complete songs (Study 2). Furthermore, participants reported experiencing more felt emotions to Studio Master music compared to the same songs heard at mp3 level.

Our covert measure of self-determined listening time did not reach significance in the present modelling due to the fact that most participants chose to listen to the complete song rather than spontaneously extending their listening time. However, in terms of the raw numbers this implicit measure of listening preference (Lindsen et al., 2011) also indicated a preference for Studio Master.

This early research into the impact of sound quality on the music listening experience should be treated with caution. At present we have no evidence regarding the source of the differences found between Studio Master and lower levels of digital resolution. Previous literature suggests possible mechanisms of effect that include, but are not limited to, a higher sampling rate (Woszczyk, 2003) and/or the presence of more inaudible higher frequencies (Kuribayashi, Yamamoto, & Nittono, 2014) in higher quality Studio Master recordings. These mechanisms should now form the basis for future controlled experiments that explore *how* enhanced sound quality might impact on the music listening experience.

We hope that this brief report stimulates future studies into the impact of high levels of musical sound quality on the listening experience that go beyond both traditional comparison stimuli (mp3 and CD) and methods (short excerpts and objective ratings). Such research would have implications not only for private music consumption but also for the ubiquitous public spaces which rely on positive explicit and implicit reactions to musical sound (North & Hargreaves, 2008).

5. ACKNOWELDGEMENTS

This research was possible thanks to an industry partnership between Goldsmiths, University of London and Linn Records. We thank Robert Cammidge, who created the music files, and Neil Parry, who installed the equipment on which the stimuli were presented. We also thank Caroline Dooley and Gilad Tiefenbrun for their support in organizing the logistics of the project. We declare that Linn Records had no influence on the design of the study, the analysis of the results or the interpretation of the data.

6. REFERENCES

Juslin, P.N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. Behavioral and Brain Sciences, *31, 559 - 575.*

Kuribayashi, R., Yamamoto, R., & Nittono, H. (2014). High-resolution music with inaudible high-frequency components produces a lagged effect on human electroencephalographic activities. Neuroreport, *Apr 9. [Epub ahead of print]*

Krause, A.E., North, A.C., & Hewitt, L.Y. (2013) Music-listening in everyday life: Devices and choice. Psychology of Music, *Published online August 14, 2013, doi: 10.1177/0305735613496860*

Krumbholz, K., Patterson, R.D., Nobbe, A., & Fastl, H. (2003). Microsecond temporal resolution in monaural hearing without spectral cues? Journal of the Acoustical Society of America, *113(5), 2790-800.*

Lindsen, J.P., Moonga, G., Shimojo, S., & Bhattacharya, J. (2011). Swayed by the music: Sampling bias towards musical preference distinguishes like from dislike conditions. Consciousness and Cognition, *20, 1781 – 1786.*

North, A & Hargreaves, D. (2008) The Social and Applied Psychology of Music. *Oxford University Press: USA.*

Olive, S. (2011). Some new evidence that teenagers may prefer accurate sound reproduction. Audio Engineering Society Convention: 131, *Paper Number: 8583.*

Pras, A., & Guastavino, C. (2010). Sampling rate discrimination: 44.1 kHz vs. 88.2 kHz. Audio Engineering Society Convention: 128, *Paper Number: 8101.*

Pras, A., Zimmerman, R., Levitin, D., & Guastavino, C. (2009). Subjective evaluation of mp3 compression for different musical genres. Audio Engineering Society Convention: 127, *Paper Number: 7879.*

Woszczyk, W. (2003) Physical and Perceptual Considerations for High-resolution Audio. Audio Engineering Society Convention: 115, *Paper Number: 5931*

Yoshikawa, S., Noge, S., Ohsu, M., Toyama, S., Yanagawa, H., & Yamamoto, T. (1995). Sound-quality evaluation of 96-kHz sampling digital audio. Audio Engineering Society Convention: 99, *Paper Number: 4112.*

Zentner, M., Grandjean, D., & Scherer, K.R. (2008). Emotions evoked by the sound of music: Characterization, classification and measurement. Emotion, *8 (4), 494 – 521.*

(RE)ORCHESTRATING TIMBRE: BEHAVIOURAL AND PSYCHOPHYSIOLOGICAL RESPONSES TO CHANGES IN ORCHESTRATION

Meghan Goodchild[1], Dominique Beauregard Cazabon[1], Jamie Webber[1], Jonathan Wild[1], Stephen Mcadams[1]
[1]CIRMMT, Schulich School of Music, McGill University, Canada
Time: 14:30

1. Background: Orchestration and timbral changes are known to influence the perception of form, contribute to the creation of musical tension, and elicit emotional responses in listeners. However, the study of timbre in music-theoretical research is underdeveloped, with few theories to explain instrumental combinations and the structural uses of timbre or their emotional impact. Given that timbre is a multidimensional phenomenon, theories of orchestration grounded in perception must account for complex interactions of many factors. One potential starting point is the often-cited correlation between spectral centroid and the perceived brightness of sounds. Orchestration treatises contain associations between brightness and positive emotional valence, as well as darkness and negative valence, but these informal accounts do not fully consider musical contexts or the relationships between the instrumentation and other musical parameters.

2. Aims: This paper directly probes the connection between orchestration and brightness and how it affects emotional response.

3. Method: Spectral centroid was modified independently through adjustments in instrumentation to explore its interaction with other musical parameters and investigate its effect on physiological responses, as well as arousal and valence ratings. Composers John Rea, Denys Bouliane, and Félix Frédéric Baril created re-orchestrations of three excerpts to produce four new versions: brightened, darkened and two hybrid combinations (dark-to-bright and bright-to-dark). Using the Digital Orchestra SIMulator (DOSIM), we created realistic acoustical renderings of the originals and the re-orchestrations. Performance timings and dynamic variations, based on real recordings, remain constant across versions. In a group-listening experiment, both behavioural (arousal and valence) and psychophysiological (biosensors) measures were collected for 108 participants (between-subjects design).

4. Results: Although the arousal ratings were mainly invariant with changes in orchestration, the valence ratings differed for brightened and darkened versions, although not always in the direction suggested by orchestration texts. Comparing responses to the hybrid recombinations (e.g., bright-to-dark vs. globally dark versions) reveals that the responses to the same passage of music are influenced by the preceding context.

5. Conclusions: Overall, we draw connections between changes in timbral brightness and emotional responses, highlight relationships between perceptual and musical dimensions, and suggest ways that these findings could be incorporated into a perceptually based theory of orchestration.

SOUNDMASS AND AUDITORY STREAMING: A PERCEPTUAL STUDY OF LIGETI'S CONTINUUM

Chelsea Douglas[1], Jason Noble[1], Stephen Mcadams[1]
[1]CIRMMT, Schulich School of Music, McGill University, Canada
Time: 15:00

1. Background: Ligeti is foremost among the composers who developed the compositional aesthetic known as "soundmass." This term has yet to be clearly defined and encompasses music of radically different characteristics — rhythmically static or kinetic, timbrally homogeneous or heterogeneous, pitched or unpitched. A complex stream may be perceived as a mass due to psychoacoustic phenomena such as masking, fusion, and spectral density, which are exploited by soundmass composers.

2. Aims: An empirical study aimed to reveal whether soundmass perception is unanimous across all listeners and which musical parameters facilitate its perception. We expect mutually interactive parameters, such as timbre, rhythmic density, and period length to influence the perception of soundmass. Ligeti's Continuum provides excellent material to study this phenomenon.

3. Method: The stimuli consisted of three MIDI versions of Continuum for harpsichord, piano, and organ presented in counter-balanced order. Participants used a slider interface on an iPad to continuously rate the degree to which they perceived soundmass. Soundmass was defined as an auditory stream that is audibly complex; the listener may perceive many simultaneous parts, but the parts do not further segregate into individual streams. A rating at one end of the scale indicates that they hear only discrete sounds that are completely segregated, a rating at the other end

indicates that they hear only a single soundmass in which all sounds are integrated, and intermediate positions indicate either that some but not all of the sounds are integrated into the mass, or that multiple simultaneous masses are perceived.

4. Results: There were no significant differences between musicians and non-musicians. Short periodicities and repetitive pitch structures increase the perception of soundmass, whereas discontinuous rhythmic structures and longer periods reduce its perception. Significant differences existed in sound-mass ratings across instruments.

5. Conclusions: Soundmass can be perceived consistently across listeners. Attack envelope and spectral composition of certain note patterns cause rating differences between instruments. The harpsichord, which has the sharpest attack, had consistent lower perceived soundmass ratings than the others, which had smoother attacks. When numerous high partials were prominent in the organ, but not the piano, soundmass ratings were significantly lower for the organ.

FINDING MEANINGS BETWEEN INSTRUMENTAL SOUND AND TAPED SOUND: LISTENING ANALYSIS OF "WATER MUSIC" BY TORU TAKEMITSU

Yuriko Hase Kojima[1]
[1]*Shobi University, Japan*
Time: 15:30

Takemitsu created "Water Music" in 1960. The piece was originally composed for 2 flutes, 1 alto flute, tape, and Noh Dance. The sound materials of the tape part are entirely created from recorded sounds of water drops that have been processed, organized and fixed. It has been considered to be one of the earliest and the most important of all the tape pieces by Japanese composers. Occasionally the tape-alone version and the chamber ensemble version are both regarded as "Water Music." However, each of them provides us with a totally different experience. It means that the tape part with electronically manipulated water sounds could be heard as one separated piece of music by itself, but once it is embedded in the sound of acoustic instruments, the tape part could be heard completely different way. Here raises a question: can listeners approach different meanings when the same music is presented in different form of performance? The ensemble version begins with solo alto flute part. This solo is smoothly connected to the tape part around 1 minute and 17 seconds. The beginning of the tape part consists of distinctively pitched attacks of the sounds of water drops that may give a listener a grasp of the pitched phrases that could be regarded as a motive. In this research, the author has tried to analyze "Water Music" focusing mainly on the ensemble version. The recording from the 2004 performance is used for the analysis, which is based on the listening analysis by solfège by the author and research participants including graduate and undergraduate composition students assisted by the use of keyboard instruments or computer soft wares to define unclear pitches and rhythms. The participants are asked to fully employ the traditional analytical knowledges and methodologies to identify groupings of pitched materials and repeated phrases or rhythmic patterns. This analytical approach hopefully could lead to a finding of relationships and meanings between the instrumental sounds and the taped sounds and compare the listening approaches for the different versions of the same piece of music.

[11D] Music and Language

LIÈGE 14:00-16:00 Friday 8 Aug 2014

TONAL AND SEGMENTAL PERCEPTION IN NATIVE CANTONESE-SPEAKING MUSICIANS, AMATEUR MUSICIANS AND NON-MUSICIANS

Ming Wai Pang[1]
[1]*State Key Laboratory of Brain and Cognitive Sciences, Department of Linguistics, The University of Hong Kong, Hong Kong, China*
Time: 14:00

1. Background and Aims: While it is established that speakers of tone languages have certain advantages in musical perception, less is known about effects of musical training on their language processing. The present study focuses on the Cantonese lexical tone and segmental perception by native Cantonese-speaking professional musicians, amateur musicians and non-musicians.

2. Method: Tone matching, judgment and segmental judgment tasks, conducted in silent reading and listening conditions, are devised to test the hypothesis that musical training improves tone and segmental (onset, rime) perception in a tone language, Cantonese, by native Cantonese-speaking individuals. Four-word sequences (in which two words are primes and two are targets or three words are primes and one is target) were presented to three groups of participants: professional musicians, amateur musicians and non-musicians in the silent reading condition, whereas four sound stimuli of Chinese characters were presented in the listening condition, and their accuracy and response time were recorded.

3. Results: Musicians, both professional and amateur, performed significantly better in tone and segmental perception than their musically naive counterparts. Moreover, the response time exhibited a contrastive pattern in the two conditions: musicians tended to respond faster in the silent reading condition, while a longer time was taken in the listening condition.

4. Conclusions: These results clearly demonstrate that musical training facilitated the perceptual processing of Cantonese tone and segmental phonemes by native Cantonese-speakers. Music-to-language transfer effects are highlighted and the non-significant differences exhibited between professional musicians and amateur musicians in five out of six tasks show that musical training need not be pursued to an advanced level for participants to gain perceptual benefits. The findings can shed light on possible forms of remedial programme development and interventions for children with language disorders such as dyslexia.

LINGUISTIC AND MUSICAL SCALING OF CANTONESE TONES

Suki S. Y. Yiu[1]
[1]*University of Hong Kong, Hong Kong, China*
Time: 14:30

ABSTRACT

Two primary areas of current research on the relationship between musical tone and linguistic tone are the intelligibility of speech tone in sung melody and the quantification of cross-linguistic tone-melody correspondence. However, there is not much work on linguistic tones independent of sung melody in musical terms. This paper attempts to express the linguistic tones in speech as musical intervals (MIs). The linguistic scaling of tones in speech can then be obtained to compare with the musical scaling of tones in sung melody. Cantonese tones are adopted as a test case.

Data was elicited from six Cantonese speakers using stimuli with a comprehensive coverage of the tonal inventory and three corner vowels. Fundamental frequencies of the tones were extracted with Praat (version 5.3.39), then time-normalized at 10% interval points across rhymes with ProsodyPro (version 4.3). The mean values of the interval points of two relatively level tones were expressed in terms of ratio, then matched with the closest MI on the musical scale. A compatible treatment of contour tones was also provided for the two rising tones.

This paper demonstrates that linguistic tones in speech and in sung melodies can be mapped with each other in terms of MIs. Both linguistically and musically, MIs can capture the tonal space, the spatial relationship among different tones in a tonal inventory, and the flexibility of different tones. The viability of MI as a mean to understand linguistic tones in speech enables a linguistic scaling of speech tones to be mapped with the musical scaling of speech tones in sung melody. The exploration of the relationship between the linguistic tones in speech and in sung melodies hopes to extend the link between the use of pitch in speech tones and music.

1. INTRODUCTION

Chao's (1956) 5-level transcription of tonal pitch variation (五度標記法) is a method developed to record the pitch value of linguistic tone with a sliding-pitchpipe. When the pitch of the pitchpipe matches the pitch of the linguistic tone, the pitch values of a linguistic tone (the starting and ending points, also a turning point between the two if any) can be notated on a staff. A tone scale with five numeric values from 1 to 5 can then be adopted to record the principal phonetic characteristic of lexical tone, i.e. fundamental frequency (F0). Tones in the Asian system, following Chao (1930), are traditionally notated with tone letter 1 as the lowest and 5 as the highest. Echoing Chao's method of notating the lexical tones via a musical mean, this paper explores how linguistic tones can be understood in terms of musical intervals (MIs) based on the phonetic data in Cantonese tones.

A MI is a perceptual distance of pitch, expressed in form of a ratio to show the distance between two notes on a scale. Figure 1 shows a chromatic scale on a piano keyboard. Different MIs are expressed in abbreviated forms: unison (U), minor (m), major (M), perfect (P), tritone (TT), and octave (O).

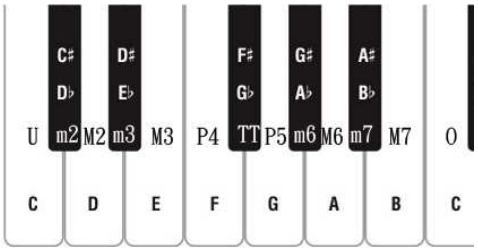


Figure 1: MIs from C on piano keyboard (Yiu 2013)

With the leftmost C as an anchor key, one step up counts as one semitone. Moving from C to C♯ involves one semitone, and its MI is a minor second. Two semitones or one whole tone is involved when moving from C to D, and the corresponding MI is a major second. The longer the distance, the more the semitones involved, and the bigger the MI. For example, the distance from C to E and from E to G♯ is the same: the MIs are both major thirds, with four semitones. As relativity of pitch is a core characteristic of phonemic tone, MI serves as a link to understand linguistic tones musically.

The MIs of linguistic tones in speech can be obtained by calculating the frequency ratios of the tone pairs in a tone inventory, a method particularly useful for describing intervals in both Western and non-Western music. The absolute pitch of one tone is expressed in the form of a ratio to another tone. The ratios can then be matched with the closest MI on the musical scale. The linguistic scaling of tones provides a platform to examine the tonal space, the spatial relationship among tones in a tonal inventory, and the flexibility of the spatial distribution of different tones in terms of MIs. Similarly, the MIs of linguistic tones in sung melody can be obtained from the frequency ratios of two adjacent tones in sung melody.

The standard Cantonese tonal inventory is composed of six phonemic tones (Matthews and Yip 1994, 2011), as listed in Table 1. The pitch trajectories of each tone are reported by numeric values. Tone values 5 to 1 denote the highest to lowest pitches. As shown, there are alternative representations for high rising tone and low rising tone. The low falling tone can also be recognized by its phonetic variant as an extra low level tone when produced in isolation (Bauer and Benedict 1997).

This paper attempts to establish a link between the use of pitch in speech, in particular phonemic tones, and the use of pitch in music. It explores how linguistic tones can be understood as musical tones in terms of MIs. The linguistic scaling of tones in speech can then be compared with the musical scaling of tones in sung melody. If the uses of linguistic tones in speech and in sung melody are related to each other via MIs, it lends support to a musical analogue to the tone inventory. Depending on the extent to which the linguistic and musical scaling of tones corresponds with each other, we can better understand the use of tones in speech and in music.

Table 1: Cantonese tonal inventory (Matthews and Yip 1994, 2011)

Tone	Tone value	Tone category
high level	55	**T1**
high rising	35/25	**T2**
mid level	33	**T3**
low falling	21	**T4**

low rising	23/13	**T5**
low level	22/21	**T6**

2. LINGUISTIC TONE AND MUSICAL TONE

The relationship between language and music is multifold. One of the relatively well studied areas is the relationship between the rhythmic structure of linguistic utterance and the general rhythmic types in music/poetry (Hayes 1985, 1995; Kager 1993; Revithiadou 2004; Hyde 2011; and many others). Concerning tone and tune, the conformity between lexical tone and melody in vocal songs has been the focus of recent works like Chan (1987), Agawu (1988), Wong and Diehl (2002), Ho (2006, 2009), Wee (2007, 2008), Sollis (2010) and Chow (2012).

Speaking and singing share a lot of similarities, for example, both involve vocalization and manipulation of pitch. However, composition of songs sometimes requires fixed pitch notation while pitch in speech tones is a relative concept in principle. Depending on the pitch of the adjacent, or even neighboring, tones in speech, the interpretation of tones may vary. If the pitch of the adjacent or neighboring tones is lower, the target tone will be perceived as a higher tone. Similarly, higher pitch of the adjacent or neighboring tones makes the target tone a lower tone (Fok Chan 1974).

For linguistic tones in sung melody, Huang (2003) and Ho (2009) observed that MIs and tonal target transitions (TTTs) are correlated with each other. TTT refers to the transition of pitch from the tonal target (i.e. offset) of a syllable to the next. When the mapping between the MI and TTT is optimal, the sung words can be perceived correctly.

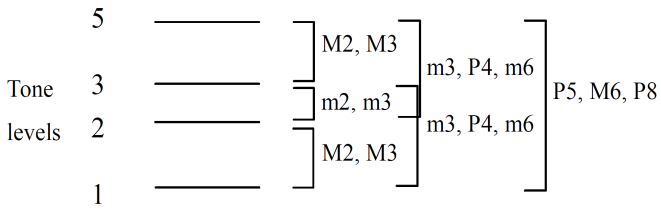


Figure 2: Optimal MIs for TTTs in Cantonese (Chow 2012)

Based on Huang (2003) and Ho (2009), Figure 2 displays the reflection symmetry in the optimal MIs of tonal transitions in vocal songs from Chow (2012). The TTTs are expressed in tonal values. Chow (2012) pointed out that the correspondence between MIs and tonal transitions in Cantonese speech can be crucial to tone perception in sung melody in Cantopop, though the overall meaning can still be deduced given the right melodic and semantic contexts even if the tone and tune do not match perfectly.

For the frequency ratios of linguistic tones in speech, Wee (2008) calculated the F0 ratios of the median of the highest and lowest tones in sung melodies for twenty Chinese languages. It questioned the musical range of an augmented fifth for the tonal scale as assumed in Chao (1930), and suggested a wider musical range. Yiu and Wee (2013) proposed MI as a possible indicator of tone merger in Cantonese. They found that only one tone pair of one of the six speakers merged in MI terms, despite F0 profiles suggestive of merger in the speech of male subjects generally. Yiu (2013) reexamined the data and found that the smallest MI was a minor second, which meant there was one semitone difference separating two tones, no matter how close those two tones appeared to be. In other words, no cases of tone merger in MI terms were reported for the subjects. Yiu (2014) further applied MIs to tones in Cantonese English in order to figure out the correspondence between the tones in the tonal inventory of Cantonese English and those in that of its substrate language, Cantonese. It showed that the tonal space of Cantonese English and Cantonese is the same (having a MI of perfect fifth) and the placement of the mid tone in Cantonese English is closer to the high tone (with a major second or minor third) than to the low tone for all speakers involved, arguing for the systemicality of the use of pitch intervals in Cantonese English.

3. METHODOLOGY

An audio-recording task was carried out to elicit the phonemic tones of Cantonese from native Cantonese speakers.

3.1. Subjects

Six Cantonese speakers balanced for gender were chosen as subjects in the project. They were all born after 1980 and raised in Hong Kong. Their age ranged from 21 to 32 years old, representing young adult speakers of Cantonese.

3.2. Data elicitation and recording procedures

Data was elicited using a wordlist which provides a comprehensive coverage of the tonal inventory, varying over three corner vowels [i], [a] and [u] in Cantonese. 18 target words were randomized with 18 fillers during elicitation.

Table 2: Target items in data elicitation

	[i]			[a]			[u]		
T1	zi1	資	‘resources’	baa1	巴	‘bus’	gu1	姑	‘aunt’
T2	zi2	指	‘to point’	gwaa2	寡	‘single’	gu2	古	‘old’
T3	zi3	志	‘gust’	baa3	霸	‘to occupy’	gu3	故	‘past’
T4	ci4	磁	‘magnet’	paa4	爬	‘to crawl’	fu4	扶	‘to carry’
T5	ci5	市	‘city’	pang5	棒	‘stick’	fu5	婦	‘woman’
T6	zi6	字	‘character’	baa6	罷	‘strike’	fu6	負	‘to minus’

Subjects were asked to produce each word three times in a row so that the basic position (initial, medial or final) of each target item in an utterance was controlled, for example, *baa1-baa1-baa1*. The list of words was recorded twice so as to take averages of the tones in each syllable position in an utterance. A total of 1296 target items (= 6 tones x 3 vowels x 3 repetitions x 2 sets x 6 speakers) were recorded. Each of the subjects contributed 108 items to the data pool. Recordings were made with Praat (version 5.3.39) (Boersma and Weenink 2013) using a sampling frequency of 22050Hz in a sound-proofed recording booth.

3.3. Method of analysis

F0 tracks of the syllables in the utterance final position were extracted with Praat (version 5.3.39) (Boersma and Weenink 2013), then time normalized at 10% interval points across the rhyme of each syllables that have been traditionally classified in the same tonal category with Praat script ProsodyPro (version 4.3) (Xu 2012).

The mean F0 of the interval points of the four relatively level tones (T1, T3, T4 and T6) were expressed in terms of ratios relative to each other. For the two rising tones (T2 and T5), each tone has two tonal points by definition. The F0 for each interval point of T2 was turned into a ratio with the corresponding interval point for T5. By doing this, both the F0 of the tonal targets, and the smallest and largest MIs would be available to decide which point(s) to be included in the calculation of MIs. Results in Yiu (2013) showed that T2 and T5 were barely distinguishable from each other until the 80% interval point, with a larger MI at a point towards the end of the contour. The same trend held for the remaining five subjects. As the farthest MI of T2 and T5 was located at the last interval point, the MI of the last point (100%) was selected for calculating MIs. The MI of each tone pair was then used to match with the closest MI on the musical scale. For the sake of accuracy, the MIs were expressed in numbers.

When the linguistic scaling of tones is ready, it was used to match the musical scaling of tones with findings in tone-melody mapping of Cantopop (Chow 2012).

4. RESULTS AND DISCUSSION

4.1. Cantonese tones on MI scale

Table 3 shows all possible combinations of ratios for speaker F2. The first column ($\bar{T}x$:$\bar{T}y$) is filled with the ratio to which each row of data corresponds. The second and third columns ($\bar{T}x$ and $\bar{T}y$) are the mean F0 of the ten interval points on the tonal contour for each tone in Hz. The column $\bar{T}x/\bar{T}y$ shows the ratios in numbers. The ratios are then matched with the actual ratio values for each MI in the next column, followed by the name of the closest MI. The last column shows the number of semitones present for the MIs classified. Let us take T1:T3 as an example. The mean F0 of T1 and T3 are 289.6875 Hz and 239.454 Hz respectively. Their ratio is 1.209783 (=289.6875 Hz/239.454 Hz), closest to a minor third in terms of MI which has a ratio of 6:5, i.e. ratio value of 1.2, with 3 semitones in between.

Table 3: MI of speaker F2

$\bar{T}x$:$\bar{T}y$	$\bar{T}x$ (Hz)	$\bar{T}y$ (Hz)	$\bar{T}x/\bar{T}y$	Closest MI ref. value	Closest MI	no. of semitone
T1:T3	289.6875	239.454	1.209783	1.2	m3	3
T1:T6	289.6875	215.8672	1.341971	1.333333	P4	5
T1:T4	289.6875	187.5193	1.544841	1.5	P5	7
T2:T5	281.0722	236.4949	1.188491	1.2	m3	3
T3:T6	239.454	215.8672	1.109265	1.125	M2	2
T3:T4	239.454	187.5193	1.276957	1.25	M3	4
T6:T4	215.8672	187.5193	1.151173	1.125	M2	2

Figure 3 shows the MIs identified for the six subjects. F/M1 to M/F3 represent fe/male speakers numbered 1 to 3 respectively.

It is found that the minimal distance between two tones in Cantonese is a minor second for the tone pair of T3:T6, with pitch levels 3-2. The maximal distance between two tones in Cantonese is a perfect fifth, which is found for the tone pair of T1:T4, with pitch levels 5-1. This coincided with the findings in tone-melody mapping, where a tonal transition from 5-1 is found to be mapped onto a musical interval of a perfect fifth (Ho 2009, Chow 2012). Figure 3 attempts to demonstrate that MI is a possible scale to portray tones in a tone inventory produced by multiple speakers, and to display the spatial relationship between different tone pairs in each speaker's tonal space. Sections 4.1, 4.2 and 4.3 will unfold the details of the tonal information available on the MI scale.

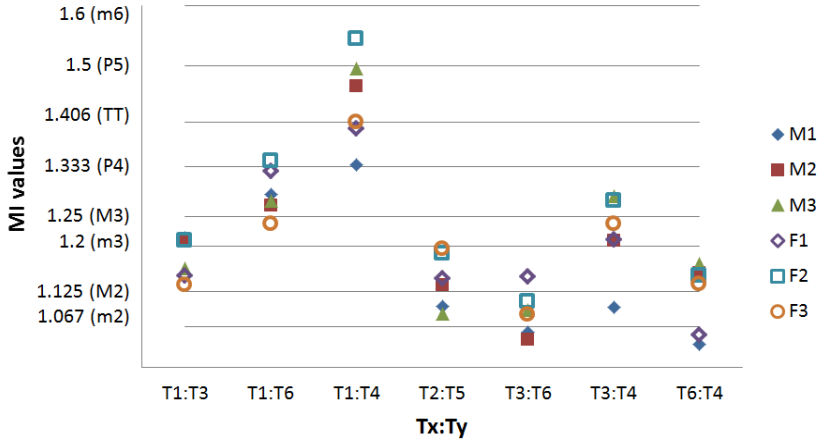


Figure 3: MIs for all subjects

4.2. Tonal space

The distance between two tones in a tone pair is reflected by the values of the MIs. The MIs of all subjects fall within the range of minor second to perfect fifth. A minor second is the shortest distance that two musical tones could have on a keyboard. A perfect fifth is the distance from C to G (cf. keyboard in Figure 1), with seven semitones in between. T3:T6 is the tonal pair with the smallest MI (minor second or major second) whereas T1:T4 is the pair with the biggest MI (major third to perfect fifth) across subjects. Other tonal pairs came within this minor second to perfect fifth range.

Speaker M1 has a relatively narrow MI range – all tones are spaced within a perfect fourth, with five semitones between the highest tone and the lowest tone. Speakers M2, M3 and F2 have the widest MI range. Their tones are spaced within a perfect fifth, with seven semitones. Speakers F1 and F3 come in the middle, having their tones spaced within a tritone, within six semitones.

Results also show that a wider tonal space allows a higher MI for the tone pair comprising the closest tones in a tone inventory. For example, the lowest MI for speaker F2 can be a major second because she is one of the three speakers who have a wider tonal space, i.e. perfect fifth as the highest MI. For the other two speakers, their minimal distance between two tones is a minor second, same as other speakers who have a narrower tonal space such as a tritone for speakers F1 and F3, and a perfect fourth for speaker M1. Considering the above, a minor second seems to be an important interval to mark the minimal distance between two tones in a tonal inventory, failing to produce it will result in confusion of tones. This leads to the next section about tone merger.

4.3. Spatial relationship among linguistic tones

The spatial relationship among different tones in a tone inventory is reflected through comparison between the MIs of different tone pairs in a tone inventory. By comparing the MIs of different tone pairs, which tones are closer to or further apart from each other can be figured out. This is especially relevant for investigating languages with a complex tone system and phonological phenomena such as tone merger.

Bauer et al. (2003) found that there were tone mergers in T2&T5, T3&T5, or T3&T6 in the speech of Cantonese speakers, resulting in various five-tone systems. With a larger sampling size, Fung (2012) found cases of full merger in T2&T5, partial merger in T3&T6, and near merger in T4&T6 in Cantonese, moving towards a four-tone system with three level tones and one rising tone. However, under a musical treatment of the F0 of the tones, Yiu and Wee (2013) showed that the tones in HKC which seemed to have merged at phonetic level still had one semitone difference separating each other. Given numerous cognitive research on music and language such as Deutsch et al. (1999) and Deutsch et al. (2004) show that the use of pitch in music and language are closely related, it is reasonable to say that the cognitive ability to process and produce the one semitone difference in music is also shared in language. In other words, tones not merged at the musical level are not merged at the phonological level likewise.

In Figure 3, the lower the MIs cluster, the closer the distance between the tones in a tone pair. Generally speaking, the sequence from the farthest pair to the closest pair of Cantonese is as follows:

T1:T4 > T1:T6 > T3:T4 > T1:T3 > T2:T5 > T6:T4 > T3:T6

As reported in Section 4.2, the smallest MI is a minor second for T3:T6. No tone merger in MI terms is found from the data.

Table 4: Optimal MIs for tonal transitions (cf. Figure 2)

Tx:Ty	TTT	Optimal MI
T1:T3 T2:T5	5-3	M2 or M3
T1:T6	5-2	m3 or P4 or m6
T1:T4	5-1	P5 or M6 or m7 or O
T3:T6	3-2	m2 or m3
T3:T4	3-1	m3 or P4 or m6
T6:T4	2-1	M2 or M3

4.4. Flexibility of linguistic tones

The flexibility of linguistic tones is reflected by the range of MIs, i.e. the extent to which the MIs of different speakers cluster for different tone pairs. The more unanimous MIs (minor second or major second) for T3:T6 confirm that the interval between these tones corresponds to minimally one semitone, as has been observed in studies of tone-melody mapping (Ho 2009, Chow 2012) as well. Meanwhile, the more flexible MIs (perfect fourth, tritone or perfect fifth) for T1:T4 indicates that a wider range of MIs are allowed.

4.5. Linguistic and musical scaling of tone

Various studies on tone-melody mapping have adopted TTTs to count the number of semitones between two linguistic tones in vocal music. Extending from that, this paper expresses the linguistic tones in speech as MIs to map with the MIs of linguistic tones in vocal music. Table 4 represents the tone values in Figure 2 as tone categories so that it would be easier to refer the MIs of musical tones identified in Chow (2012) to the MIs of linguistic tones identified in this paper.

Table 5 gives a categorical representation of the MIs of linguistic tone in Figure 3. The shaded boxes are the MIs of linguistic tones which match the MIs of musical tones.

Table 5: MIs of all subjects (cf. Figure 3)

Speaker / Tx:Ty	M1	M2	M3	F1	F2	F3
T1:T3	m3			M2	m3	M2
T1:T6	M3			P4		M3
T1:T4	P4	P5		TT	P5	TT
T2:T5	M2		m2	M2	m3	
T3:T6	m2			M2		m2
T3:T4	m2	m3	M3	m3	M3	
T6:T4	m2	M2	m3	m2	M2	

Though the MIs of linguistic one for individual speakers do not seem to map perfectly with the MIs of musical tone, it is clear from the mapping that the musical scaling is wider than the linguistic scaling by allowing more flexible MIs for the same tone pair. Whenever the linguistic tones in speech and music do not match, it is often because the musical scaling has MIs higher than what we found in speech, without taking care of the lower MIs of some tone pairs for certain speakers. This could be due to musical performance generally involving a broader (in particular higher) range of pitch use than normal speech.

Both linguistically and musically, the near unanimity of MIs (minor second or major second) for T3:T6 confirms that the interval between these tones corresponds to a semitone whereas the more flexible MIs (major third, perfect fourth, tritone or perfect fifth) for T1:T4 reflects that a wider range of MIs are allowed.

The linguistic tones in speech and music do not seem to map better for any particular tone pair or speaker. The reason for that may be the flexible nature of the MI scale since the MIs are approximate intervals based on the distance between two linguistic tones. Still, the use of a flexible MI scale goes along with the fact that humans do not produce MIs with absolute accuracy in natural speech.

4.6. MI as a tone scale

Since the principle phonetic characteristic of tone is pitch, carefully choosing a scale to represent tones that best represents the perceptual image of the production data is important.

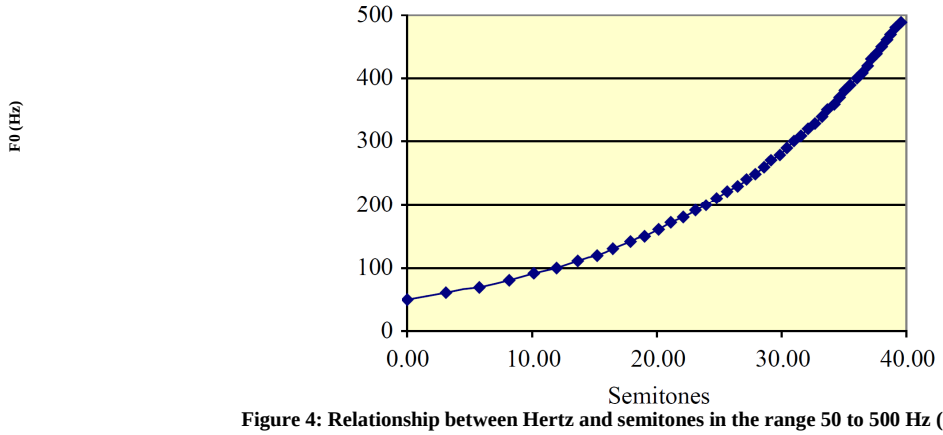


Figure 4: Relationship between Hertz and semitones in the range 50 to 500 Hz (Nolan 2003)

Our perception and production of pitch is generally non-linear, unlike the Hertz scale. There are a handful of psycho-acoustic scales available to provide steps corresponding to equal perceptual intervals. The best known among them is the musical semitone scale, which divides each musical octave into twelve equal increments as semitones. Nolan (2003) demonstrated that a logarithmic (semitone) or near logarithmic (ERB-

rate) scale best models speakers' intuitions about intonational equivalence. Later works related to linguistic scaling of pitch such as Kügler (2009) and Rialland and Somé (2011) also adopted the musical semitone scale.

Similar to the semitone scale, the proposed MI scale also makes use of semitone differences. However, the semitone scale may requires the semitone values of different speech tones transformed by an equation built-in with some reference values whereas the MI scale obtains the MI values from F0 ratios of different tone pairs. The MI scale is more superior in the sense that it captures the relativity nature of phonemic tones in speech and expresses the distance of tones in different tone pairs as MIs straightforwardly. This is particularly useful for comparing tones within and across tone systems based on F0 data of multiple speakers.

MIs allow tonal analysis be done without committing to specific tuning systems which divide an octave slightly differently, avoiding small yet possibly significant effects of such differences on the interpretation of the spatial relationship of different tones in a tone inventory. Comparison of linguistic tones can be made through MI ratios ranging from 1 to 2 in an octave. Switching from one tuning system to another can be done by changing the ratios which the MIs are based on. This paper adopts the just intonation tuning system as provided by 5-limit tuning, where the main intervals and just intervals are found to sound well-tuned to most people and are commonly used in musical cultures worldwide. MI ratios can be divided into smaller logarithmic units like semitones by $\log_2 MI * 12\log_2 \mathrm{MI}*12$, and so as cents by $\log_2 MI * 12\log_2 \mathrm{MI}*12 * 100\log_2 \mathrm{MI}*12*100$.

MI as a tone scale also makes comparison between the linguistic scaling of tones in speech and the musical scaling of tones in sung melody possible. As shown in section 4.5, the tonal space, the spatial relationship among different tones, and the flexibility of tones are visible on both linguistic scale and the musical scale. The MI scale plays a role in understanding the use of pitch in language and music by linking linguistic tone and musical tone.

5. CONCLUSIONS

This paper has demonstrated that linguistic tones in speech can be expressed in terms of MIs based on data from Cantonese. Both linguistically and musically, MI as a tone scale can capture the tonal space, the spatial relationship among different tones in a tonal inventory, and the flexibility of tones. The viability of MI as a mean to understand linguistic tones in speech enables a linguistic scaling of speech tones to be mapped with the musical scaling of speech tones. MIs allow a straightforward transformation and representation of linguistic tones in musical terms. A musical analysis of tones with MIs can be done without committing to specific tuning systems which divide an octave slightly differently. Depending on the musical culture of the language speaker, an appropriate tuning system can be chosen to divide the MIs accordingly. A larger language sample with the corresponding musical traditions needs to be studied to assure the applicability of the MI scale.

On top of the above, the bigger question that this paper leads to concerns precision of the placement of different tones in a tonal space to achieve phonemic contrast, especially for languages with a complex tone system. Even with advanced recording and analytical tools, to the surprise of many, this is no easy question. Founded on the shared cognitive ability and similar manipulation of pitch use in language and music, this paper hopes to shed some light on the scientific pursuit of a more comprehensive representation and analysis of tone.

6. ACKNOWLEDGEMENTS

I thank Stephen Matthews and Diana Archangeli for their insightful discussion and encouragement along the way. I also thank Lianhee Wee, Jon Yip, Richard Chim, 'lab rats' Winnie Cheung and Queenie Chan, and the six subjects. I also appreciate the comments from the audience at the International Conference on Phonetics of the Languages in China 2013 and Speech Prosody 7 on two relevant papers.

7. REFERENCES

Agawu, V. Kofi. (1988). Tone and Tune: The Evidence for Northern Ewe Music. Africa: Journal of the International African Institute 58, *no. 2, 127-146.*

Bauer, Robert S. and Paul K. Benedict. (1997). Cantonese Tones. In Modern Cantonese Phonology, *109-250. Berlin; New York: Mouton de Gruyter.*

Bauer, Robert, Kwan-Hin Cheung, Pak-Man Cheung. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. Language Variation and Change *15:2, 211-225.*

Chan, Marjorie K.M. (1987). Tone and Melody Interaction in Cantonese and Mandarin songs. UCLA Working Papers in Phonetics *68:132-169.*

Chao, Yuen-Ren. (1930). A System of Tone Letters. La Maître phonétique *45:24-27.*

Chao, Yuen-Ren. (1956). 现代吴语的研究.科學出版社.3,4,74.

Chow, Man Ying. (2012). Singing the Right Tones of the Words: the Principles and Poetics of Tone-melody mapping in Cantopop. *MPhil dissertation: The University of Hong Kong.*

*Deutsch, Diana, Trevor Henthorn, and Mark Dolson. (1999). Tone language speakers possess absolute pitch. Invited Lay language paper presented at the*138th meeting of the Acoustical Society of America*. October, Columbus.*

Deutsch, Diana, Trevor Henthorn, Elizabeth Marvin, and Hongshuai Xu. (2004). Tone language and absolute pitch: Prevalence among American and Chinese conservatory students. Journal of Acoustical Society of America *116, 2580.*

Fok Chan, Yuen-yuen. (1974). A Perceptual Study of Tones in Cantonese. *Hong Kong: The University of Hong Kong.*

Fung, Suk-Yee. (2012). Tone mergers in contemporary Hong Kong Cantonese. Departmental Seminar, The University of Hong Kong.

Ho, Vincie W.S. (2006). The Tone-Melody Interface of Popular Songs Written in Tone Languages. In Proceedings of ICMPC9, *University of Bologna, Italy, 1414-1422.*

Ho, Vincie W.S. (2009). Fine-Tuning Tone-Melody Constraints through the Investigation of Mismatches in Cantonese Pop Music. Hong Kong: City University of Hong Kong.

Huang, Zhihua. (黃志華) (2003). 粵語歌詞創作談 [On Writing Cantonese Lyrics]. Hong Kong: Joint Publishing.

Kügler, Frank (2009) A model for the quantification of pitch accent realisation. In: Kügler, F., Féry, C. & van de Vijver, R. (eds) Variation and Gradience in Phonetics and Phonology. *Berlin: Mouton de Gruyter, 405-424.*

Matthews, Stephen, and Virginia Yip. (1994, 2011). Cantonese: A Comprehensive Grammar. *London: Routledge.*

Nolan, Francis. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In Proceedings of the 15th international congress of phonetic sciences, *771-774.*

Rialland, Anne and Penou-Achille Somé (2011). Downstep and linguistic scaling in Dagara-Wulé. In: Goldsmith, John A, Elizabeth Hume and Leo Wetzels. (eds) Tones and Features. Phonetic and Phonological Perspectives. *Berlin, Boston: De Gruyter Mouton, 108-134.*

Sollis, Michael. (2010). Tune-Tone Relationships in Sung Duna Pikono. Australian Journal of Linguistics*30,no.1:67-80.*

Wee, Lian-Hee. (2007). Unraveling the Relation between Mandarin Tones and Musical Melody. Journal of Chinese Linguistics *35, 1, 128-144.*

Wee, Lian-Hee. (2008). Inquiry into the Musical Nature of Linguistic Tone. In Hsiao, Yuchau, Hui-chuan Hsu, L.H. Wee and Dah-An Ho (eds) Interfaces in Chinese Phonology. *Taiwan: Institute of Linguistics, Academia Sinica, 139-160.*

Wee, Lian Hee, Queenie K.Y. Chan, Winnie H.Y. Cheung and Suki S.Y. Yiu. (2013). Tone samples of Cantonese and English in Hong Kong. *Phonology Laboratory, Hong Kong Baptist University.*

Wong, Patrick C. M., and Randy L. Diehl. (2002). How Can the Lyrics of a Song in a Tone Language Be Understood? Psychology of Music *30, no. 2: 202-209.*

Xu, Yi. (2012). Praat script ProsodyPro (version 3.5)

Yiu, Suki S.Y. (2013). Cantonese Tones and Musical Intervals. In W.S. Lee (Ed.) Proceedings of the International Conference on Phonetics of the Languages in China 2013 (ICPLC 2013), *Hong Kong: the Organizers of ICPLC 2013 at the Department of Chinese, Translation and Linguistics, City University of Hong Kong, 155-158.*

Yiu, Suki S.Y. (2014). Musical Intervals of Tones in Cantonese English. In N. Campbell, Gibbon, D., and Hirst, D. (Eds.) Proceedings of the 7th International Conference on Speech Prosody (SP 7), *587-591*.

Yiu, Suki S.Y. and Lian-Hee Wee. (2013). Hong Kong Cantonese Tones: Replication Study. Oral presentation at the 13th Workshop on Cantonese (WOC 13). *Hong Kong: Hong Kong Baptist University, March 16.*

INVESTIGATING THE INTEGRATION OF LONG-DISTANCE DEPENDENCIES IN LANGUAGE AND MUSIC

M. Paula Roncaglia-Denissen[1,2], Fleur Bouwer[1,2], Henkjan Honing[1,2]
[1]Institute for Logic, Language and Computation, Universiteit Van Amsterdam, Netherlands, [2]Amsterdam Brain and Cognition, Universiteit van Amsterdam, Netherlands
Time: 15:00

1. Background: Despite differences in functionality and domain-specific elements, music and language may share a common processor to integrate their syntactic structures (for an alternative view, see Peretz & Coltheart, 2003). This is known as the Shared Syntactic Integration Resources Hypothesis (SSIRH; Patel, 1998, 2003,) and is supported by quite some evidence (Maidhof & Koelsch, 2010; Slevc, Rosenberg, & Patel, 2009). Nevertheless, it might actually be shared attentional resources and not so much a shared syntactic parser that account for these findings (Perruchet & Poulin-Charronnat, 2013). Due to greater memory and attentional load, long-distance dependencies provide a good candidate to further investigate this matter. Despite research on long-distance dependencies processing in language (cf. Frazier, 2002) and music (Koelsch, Rohrmeier, Torrecuso, & Jentschke, 2013), no study has addressed this issue in these domains combined (cf. Arbib, 2013)

2. Aims: This research investigates if language and music make use 1) of a common syntactic processor or 2) shared attentional resources during the processing of long-distance dependencies.

3. Method: Participants are visually presented with sentences containing a syntactic violation of long-distance dependencies while hearing melodies presenting violation of long-distance tonal expectations. Sentences contain one main clause followed by a relative clause agreeing with or violating number agreement in the main clause. Melodic phrases conform the Western tonal hierarchy and present three parts. In part one, a clear key context is established, while an ambiguous key is provided in part two. In part three, a single note either conforms previously presented keys or violates the key of part one only. Participants will either judge the correctness of linguistic sentences (language task), the adequacy of melodic closure (music task) or both (attention-divided task).

4. Results and Conclusion: If language and music share a syntactic processor, slower reaction times and lower accuracy rates are expected in response to the double violation during language and music processing in all three tasks. If, however, worse responses are found in the attention-divided task compared to the others with no interaction between language and music processing, this might indicate that previously reported interactions between these domains could result from shared attentional resources.

CONSTITUTION OF MUSICAL MEANING: A PRAGMATIC APPROACH TO CONSTITUTIONS OF MUSICAL SIGN AND ITS UTTERANCE MEANING

Son-Hwa Chang[1]
[1]University of Cologne, Germany
Time: 15:00

ABSTRACT

This paper proposes a pragmatic approach to musical meaning from the perspective of narrative understanding. Comparing musical understanding to understanding of fictional narrative this paper examines the possibility of the constitution of purely musical signs and their meaning in the unfolding of a musical work. It is proposed that identification of musical sign (initially without any specified meaning) is achieved by phrasing of music by means of plot-like structure inherent in a musical phrase. Then the role of repetition and variation in the constitution of musical sign in the course of interpretation of music is discussed. Finally, constitution of musical meaning carried out via exemplification within a musical work is discussed.

1. INTRODUCTION: MUSICAL PRAGMATICS

Musical meaning is a branch of aesthetics and important field of music research in musicology, psychology, philosophy, cognitive neuroscience of music, the affective sciences, and cognitive poetics. It has been investigated predominantly as semantics in relation to musical syntax and emotion (e.g. (Koelsch, 2011), (Patel, 2008)). However semantic investigations of musical meaning has their limitation in coping with the main characteristics of musical meaning like intra- and intersubjective variability (Faltin 1973, 1978), ambiguity (Cross 2005), fluidity (Kühl

2007), and ineffability (Raffman 1993). To cope with such specific challenges an approach to musical meaning from the viewpoint of pragmatics is proposed as a complement to musical semantics for the investigation of musical meaning.

Charles William Morris (1938) introduced **pragmatics** as one of three components of his semiotic trichotomy besides syntactics, semantics. It is "that branch of semiotic which studies the origin, the uses, and the effects of signs." (Morris 1946: 352) According to him it is an all-embracing component that presupposes both syntactics and semantics, and deals "with the biotic aspects of semiosis, that is, with all the psychological, biological, and sociological phenomena which occur in the functioning of signs." (Morris 1938: 30) After its grounding pragmatics is mainly understood as a part of linguistic discipline that concerns essentially with fine-grained components of meaning conveyed in an actual situation of language usage and in connection with varying contexts in which utterances are embedded (Levinson 1983). It is usually concerned with linguistic phenomena related to the contexts of utterance situations, such as prosody, intonation, and deictic terms, etc., cognitive capacities underlying linguistic behaviours, e.g. implicature, and linguistic activity and its socio-cultural effect, like speech act.

Generally pragmatics can be characterised as an investigation to communicative sense—i.e. what is meant and understood by uttering in an actual communicative interaction (Löbner 1996). It strives to figure out the variable and constitutive aspects of meaning of utterances related to the varying context of communicative interaction. For this it examines diverse modes of signing in communication, focussing especially on the ostensive mode in which the content of communication is rather shown than referred to by signs. (Sperber and Wilson 2004) The showing mode of signing has several different names, like “presentational sign” (Langer 1979) and “exemplification” (Goodman 1968). In the showing mode of signing the material features of used signs contribute to meaning. The material features are usually distinguished as the form component of a sign from the meaning component This showing mode provides the possibility to convey indefinitely dense and fine-grained information that cannot be explicitly told.

As an effect of vivid interdisciplinary comparative research on the evolution of language and music pragmatic investigations to musical meaning is recently growing in music research. In language and music evolution research the interplay between syntax, phonetic/phonology, and semantics is discussed in connection to the emergence of generative sign/symbol systems.(e.g. Fitch 2010, Arbib 2013) Further important topics are among others the role of sound effects and segmentation in constituting signs (e.g. Verhoef, Kirby, & Padden 2011) and their meaning (e.g. Hurford 2007). Pragmatic investigations to musical meaning in music research are to be found in studies concerning the role of sound effects for musical meaning in evolutionary context (Cross 2005, Cross and Woodruf 2009), the relevance of repetitiveness for learning and memory constitution (Margulis 2014, Richman 2000), the role of sound pattern in voice in meaning constitution and the musicality of temporal organisation of turn-taking (Gratier and Trevarthen 2007, Levinson 2013). In addition research on musical narrativity can be understood as a branch of musical pragmatics (Cone 1974, Newcomb 1984a, 1984b, Gratier and Trevarthen 2008, and Trevarthen 2008).

Musical pragmatics offers new perspectives for interdisciplinary investigations on meaning constitution, understanding, and aesthetics in general. Especially investigations of narrativity in music seem to provide promising contributions to cognitive aesthetics (Schellekens and Goldie 2011, Turner 2006) in connection with the cognitive capacity for fiction and fantasy. This paper focuses on the pragmatics of musical narrative and its role in the constitution of musical meaning.

2. STORYTELLING, NARRATIVE UNDERSTANDING, AND MUSICAL NARRATIVE

Storytelling is the act of representing or materializing a story. In a story several events and entities are related to each other in a coherent way so that it is understandable as a whole. A story is a mental construct that needs to be materialized in a certain way. Storytelling is the act of materialization of a story. The product of storytelling is a narrative, i.e. text in materially realised form,that may evoke a story in mind through reading or listening. (Ryan 2008): A narrative is a materialized representation of a story.

In storytelling events are not just put in temporal order of their happening. Depending on logical comprehensibility and the need of dramatic effect events are arranged in such a manner that the resulting sequential representation attracts curiosity of the audience and relieve their understanding of the relations between events and thus the course of development of story. This implies that emotional involvement of diverse kinds is essential for understanding a narrative.

A coherent dramatic structure of a narrative called “plot” is a result of such an rearrangement. A plot depends on several factors of affective nature, such as a general pattern of predicament, i.e. its emergence, development, and resolution, and dynamic changes of “the inner life of the experiencing subjects” (Cone 1974, 86) called “persóna”. The affective states involved in a plot assist narrative understanding.

In music research conceiving of musical understanding as a kind of narrative understanding has been mainly related to the previously mentioned emotional involvement that accompanies narrative understanding. Edward D. Cone (1974), for example, claims that music can be conceived of as "an expression of the inner life of the experiencing subject." (Cone 1974, 86) His investigation to narrative aspects of musical understanding focuses on the identification of musical persona, i.e. the experiencing subject in a musical work and the expression of their inner life in music. Especially he considered the possibility taking certain instruments or groups of instruments as voices of musical persónas, which, then, enables a complex and differentiated understanding of music. Anthony Newcomb (1984) examined the "plot" or "plot-archetype" underlying musical structure and its development in a whole work. Musical structure itself is considered as logically and dynamically comprehensible. He claims that, just like that several plot archetypes that are underlying novels and tales as basic structure, at the basis of absolute music there is “an evolving pattern of mental states” (Newcomb 1984b, 234). According to Maya Gratier and Colwyn Trevarthen (2008) musical structure as essentially narrative and music’s inherently sequential organization provides modulated tension and energy.

These investigations so far focused mainly on structural and dynamic aspects of music but—in particular in the case of instrumental music—largely neglected the important aspect of conveying the content of a story by means of music. That led to skepticism about musical narrative (e.g. Wolf 2008). Narrative understanding is a complex mixture of every component of storytelling. Other components can surely complement the understanding of a story but not replace it. Without a story the narrative content itself remains empty. In the study of musical narrative one of the most important question is therefore whether music can have an extramusical meaning that would make the understanding of a concrete story possible.

3. MEANING AND STORY CONSTITUTION

Investigating meaning constitution in music from the viewpoint of pragmatics might be one promising way to give new insights into the problem of extramusical meaning. There are already some examples of this kind of effort. Cross (2005, See also Cross & Woodruff 2009) is

concerned with the constitution of intersubjectively sharable basis for musical meaning in actual musically mediated social interaction. Embodied music cognition (Leman 2008) investigates formation of embodied musical meaning in the course of a collectively extended action-perception cycle in the course of music history. Gratier and Trevarthen (2008) considering not only musical form but also the content of music as essentially narrative also emphasize the importance of historically mediated constitution of cultural meaning.

The other way is to reconsider the nature of narrative understanding: To what extent is an exact referential meaning needed for story understanding? Zwaan and colleagues (Zwaan, Magliano, & Graesser 1995; see also Zwaan 1999) showed that a situation model of the supposed situation which creates the context a story is embedded in, rather than the exact knowledge of each elements indicated in a story, makes a story understandable. Furthermore, it has been proposed that human understanding itself is basically narratively organised (Turner 1996). In such a case even implicit meaning inherent in the form component of a musical work would be sufficient for activating narrative imagination that contributes to the understanding of musical narrative. If a situation model is constituted, musical signs can be taken as referring to imaginative objects constitutive for a story and so music gains concrete referential meaning in relation to that established context.

The aim of this paper is to discuss the constitution of musical sign/meaning in the course of the production-perception cycle of a musical work. This kind of extended interaction mediated by a musical work is a special case of interactive communication in which musical signs are constituted in the unfolding of the work and at the same time used. In particular, the listening process of a musical work following its all nuanced unfolding is considered as a process of constitution of a fictional story. The musical signs used in a work obtain an intersubjectively shared meaning on the basis of the fictional story which is constituted while listening. The remainder of this paper concentrates on the question how this story constitution is proceeding.

4. THE CONSTITUTION OF MUSICAL SIGN

As already indicated, the most obvious problem of considering musical understanding as a narrative understanding is, the fact, that in music there is no preexisting sign with conventionalized referential meaning. This is the reason why musical meaning is often considered to be abstract. In this section I propose that musical discourse is formed within a framework of its structural unfolding during the process of listening to a musical work. In the course of formation it gains a meaning that is though not always clear and distinct but distinguishable to something else. This distinctness provides the possibility to constitute an utterance meaning, i.e. the meaning of an utterance in a given context. The constitution process consists of three parts: unit identification, sign- and meaning-constitution, and division of feature dimensions.

4.1. Unit Identification

The first concern to discuss is how a musical sign can be identified at all, if there is no preexisting sign. Should then every distinguishable events beginning from a single note to a whole musical work be taken as musical signs? In their Generative Theory of Tonal Music, Lerdahl and Jackendoff (1983) proposed segmentation rules that serve to figure out possible boundaries for groups as musically identifiable units. They have shown that musical grouping structure is hierarchically organised. Then the previous question could be put in a more precise form: At which level may musically identifiable units function as musical signs with extramusical meaning? This leads to the question which level of units are most probably able to carry an extramusical meaning.

I propose plot-like structure inherent in a musical phrase as a most probable candidate for a musical sign with an extramusical meaning. A plot is a structure of tension that can automatically be identified as a unit. Gratier and Trevarthen consider (2008) plots as underlying the temporal organisation of vocal interchange in affective social interaction between infants and caregiver and therefore regulates the turn-taking. It has been shown that the sound length of a turn is intuitively recognized based on the seemingly innate knowledge about sound contour (Malloch & Trevathen 2008, see also Levinson 2013).

Plots are furthermore considered as meaningful. Defining narrative as “dynamic, temporally organised action or praxis” (Gratier and Trevarthen 2008: 125), Gratier and Treverthen claim (2008) that a plot provides an energetic contour of narrative tension that “creates dramatic moments that stand out and that are memorable as shared emotional experience” (Gratier and Trevarthen 2008, 127) And this shared emotional experience provides then the basis for a constitution of practical and social meaning. It seems then that plots, the automatically identifiable and inherently meaningful units, could serve as the basic unit from which sign and meaning constitution proceed.

So aesthetic units with plot-like structure and a natural feeling of closure form the starting point for the constitution of extramusical meaning.

4.2. Sign- and Meaning-Constitution via Repetition and Exemplification

Unit identification establishes the basis for sign and meaning constitution. In order to function as signs, identified units need to be recognized by a listener as signs—they have to be taken as standing for something, i.e. their sign function needs to be realized—and they have to carry meaning. Furthermore, to be a sign with intersubjectively sharable extramusical meaning, the essential feature of an identified unit has to be figured out so that its utterances can be recognized. Then the following questions arise: How does a unit come to be taken as a sign? How becomes a unit an utterance of a sign? Where does its meaning come from?

The answer to these questions lies in an attentional shift in the listeners mind from the flow of ordinary actions to the identified unit and its perceptual features. This shift leads to a discovery process directed to the unit’s subtle and attractive properties and interesting associations accompanied by pleasure. This attention-discovery state is usually considered as being aesthetic and mainly discussed in aesthetic context. For example, Ramachandran and Hirstein (1999) proposed eight brain mechanisms for aesthetic perception of visual art that enable such an attentional shift and the discovery of pleasurable features contained in given perceptual information. The most prominent is a peak shift effect—This effect has been observed in animal discrimination learning in which animals are learning to response preferably to rewarded stimulus, especially to stimuli which have more distinctive differentiating features.—So animals are able to learn to figure out the perceptual essence of a given stimulus. This preferred response to supernormal stimuli indicates that animals figure out the essential features of concern. In his commentary to Ramachandran and Hirstein Colin Martindale (1999) points out that the peak shift effect occurs only then, when the differences between the compared stimuli are subtle. This indicates that it is the subtle variation in repeated exposure to objects that makes the detection of the essential features possible.

Once essential features of an identified unit have been figured out, it can be recognized as one of the utterances of a sign. However, it presupposes that it has been taken as standing for something. How does its sign-function recognized? The answer for this question lies also in an attentional shift as well as in the mechanisms of repetition and variation. In her discussion about the functions of repetition Elizabeth Helmuth

Margulis (2014) pointed out that repetition causes an attentional shift from conveyed information to something more implicit in the repeated unit itself. This shift leads the listener to pay more attention to the uttered sign itself seeking in it implicit meaning, which she identifies with "ever-richer associations and patterns" (Margulis 2014, 13) conveyed by the utterance. Margulis (2014) terms this kind of attentional shift "aesthetic orientation" and claims that it is characterized by its "perceptual openness, a willingness to notice and believe in connections and meanings that may not be instantly apparent." The listener's search for and believe in something more implicit what makes an identified and repeated unit a sign. Although there is no "real" meaning yet, such a search for meaning opens up the possibility to fill in the empty place of meaning with ever-richer associations and patterns implied in an utterance.

Nelson Goodman (1976) investigates such search for not yet existing meaning that creates new meaning. In what he terms "self-exemplification" an identified unit thing identified becomes a symbol by standing for itself while attracting all attention to itself so that a search for meaning can begin. A self-exemplifying symbol is in the showing mode of signing which he calls "exemplification". In exemplification one of the objects having certain properties is taken as a sample to refer to that properties (Goodman 1976). In self-exemplification a sample refers to all of its properties prompting the interpreter to discover all its interesting properties in their nuanced and subtle form. However, if only the mere features of an exemplifying sign constitute the sign's meaning, then, in the case of music, it is at best an intramusical meaning. But what about extramusical meaning? Where does extramusical meaning come from?

In order to discuss about the constitution of extramusical meaning, we have to turn at first to the division of feature dimensions and the idea of the predication by sample.

4.3. Division of feature dimensions

In discussing null denotation—i.e. reference to nonexistent entities, like fictional entities—Goodman (1976) suggests that one doesn't have to know what a symbol means to use a symbol. Rather, one learns the meaning of a symbol from its usage, especially from the combinations of symbols. For example, we can learn what the word "ice" mean, by learning when it is used in different compound terms, like "ice cream", or "iceberg". Likewise we don't have to know what a unicorn is to learn what a unicorn picture is. We learn from different unicorn pictures and descriptions (hereafter in short: representations) about unicorns. This is the way how non-existing, fictional entities are becoming to exist. Although there is no entity to refer to, by pretending to refer to something one can create an entity referred to in a fictional world of imagination (Goodman 1984). This explains how referential meaning is created.

Goodman even goes a step further: If one is able to identify a sign pretending to refer to something, he/she can learn from the sign about the fictional entities created by it. By learning to categorize some objects as unicorn representations, one can learn from unicorn representations what an unicorn like/ how an unicorn looks like. (Goodman 1976) This is possible insofar as such representations not only are pretending to refer to something, i.e. a fictional entity *x*, but also depicts or/describes it "as" something, e.g. a unicorn, in a certain way, i.e. having certain properties (cf. figure 1 in the case of music). According to Goodman every representation represents its referent as something putting some properties of it in foreground. So while a fictional representation (or a part of it) is referring as a whole to a fictional entity, some details of the representation can be taken as depicting/describing properties of the entity. As a result, by knowing what the representation is about, one can learn about the represented non-existing entity from the way how it is represented.

Now here is an interesting constellation: A new created fictional entity has no properties, because a non-existing entity cannot have any property. However, since the fictional entity is represented as something, the entity that initially had no features has received features as its properties: The features ascribed to the entity. In other words, a representation is a case of predication by which a new entity with properties and affective states can be constituted. Through this predication a fictional entity becomes a concrete being in the world of narrative imagination that is evoked by a representation.

A fictional representation brings out a new entity *x* as its referent via predication. It creates its own meaning that lies beyond the range of its own features or mere associations evoked by the features. The listener's search for something more implicit discussed in previous section is then not limited to the perceptual features contained in the sign. It is more active and creative in nature: Filling in the initially empty place of meaning (i.e. the new created fictional entity *x* without any feature) with subtle and fine-grained perceptual features contained in the utterance, listeners use their narrative imagination to discover in some of these features what could possibly possessed by that entity and ascribe them to that entity, so that during the course of listening the entity itself is created in a concrete way. This constitution of an "aesthetic object" (Ingarden 1968) is the ultimate aim of a narrative understanding as a result of aesthetic perception.

So far meaning constitution via predication has been discussed in general. How, then, does musical meaning constitution proceed? To answer this question one has to consider what is meant in musical sign by "part". It was already proposed that a sign representing as a whole is a plot-like structure in music. In the case of pictorial or verbal representation parts of a representing sign can be clearly distinguished. In an unicorn picture, for example, it is easily recognizable which part forms the fictional animal's corpus and which the horn. hat could serve in music as part of a representing musical sign that can be taken as depicting/describing properties of the entity *x*? Can a distinctive part of music function as a predicate?

In Prokofieff's symphonic tale "Peter and the Wolf" an narrator introduces each character of the tale as represented by different instruments or groups of instruments, e.g. the bird is represented by a flute, the duck by an oboe, and Peter by the group of string instruments. While being introduced, each instrument representing a character is playing a short characteristic musical passage: Musical samples for each character are given. But not every musical passage these instruments are playing represents the characters. However, knowing that certain instruments are representing a certain character, each time when these instruments are playing a solo or the musical passage introduced at the beginning of the work, one can identify the occurrence of the character. These musical utterances have several musical properties, such as certain rhythmic and melodic patterns, tempo, dynamic, a specific length of phrasing, different articulations, timbre and so on. Some of these properties are varying in each occurrence and some remain rather constant.

It seems then it is the difference in the dimensions of a musical features that are separated into parts to serve as musical predicates. Which dimensions are to be taken as characteristic features for sign discrimination and which for predication differs by each character and musical unfolding: This means, it remains the achievement of the listener to figure out the exemplifying properties to hear a story in music. The following figure shows the process of musical predication.

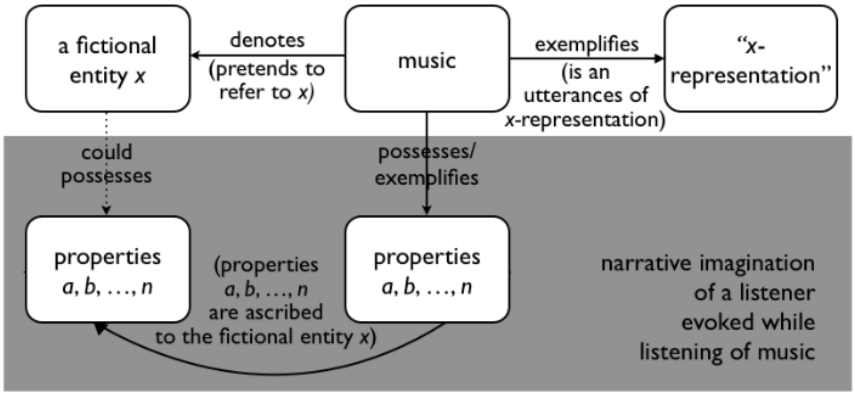


Figure 1: Musical predication. During the listening of the unfolding of a musical work a listener discovers properties of music that could be shared by the represented fictional entity and ascribes these properties to the entity.

5. CONCLUSION

It was argued that investigating music as pragmatics and from the viewpoint of storytelling and narrative understanding, i.e. as musical narrative, provides a promising approach to integrate different investigations to musical meaning related to emotion, language and music processing focusing on the most interesting characteristic of musical meaning, namely its complex, dynamic and creative aspects. The listener's narrative imagination induced by music brings extramusical meaning out and makes it concrete during the course of listening so that it gains its existence in her imaginary world. A plot-structure and a situation model which sets up a context for storytelling and narrative understanding provide the frame to guide this specific narrative imagination. Therefore despite the obvious absence of conventionalized referential signs music is able to tell stories by creating signs with extramusical meaning during the unfolding of a work in the process of listening.

Repetition is proposed as a main mechanism for sign- and meaning-constitution. Furthermore, repetition plays an important role in providing coherence for musical works by keeping a topic-structure constant. Musical coherence needs further investigation and will be discussed elsewhere.. The same holds about the role of voice in music in giving music coherence and attracting emotional involvement more effectively. Gratier (2008) gives interesting insights into this topic.

An approach to the riddle of (musical) meaning based on pragmatics will have implications for the future investigations in Cognitive Poetics and psychological research on film music as well as music theory and music cognition.

6. REFERENCES

Arbib, M.A. (ed.). (2013). Language, music, and the brain: A mysterious relationship, *Cambridge, MA: The MIT Press.*

Cone, E.T. (1974). The composer's Vvice, *Berkeley, Los Angeles & London: University of California Press.*

Cross, I. (2005). Music and meaning, ambiguity and evolution. In D. Miell, R. MacDonald & D. Hargreaves (eds.), Musical communication *(pp. 27-43), Oxford: Oxford University Press.*

Cross, I. & Woodruff, G. E. (2009). Music as a communicative medium. In R. Botha & C. Knight (eds.), The prehistory of language *(pp. 113-144). Oxford: Oxford University Press.*

Faltin, P. (1973). Die Bedeutung von Musik als Ergebnis sozio-kultureller Prozesse: Zu einigen Schwierigkeiten bei der Betrachtung der Musik als kommunikatives Phänomen, Die Musikforschung, *26(4), 435-445.*

Faltin, P. (1978). Musikalische Bedeutung. Grenzen und Möglichkeiten einer semiotischen Ästhetik, International Review of the Aesthetics and Sociology of Music, *9(1), 5-32.*

Fitch, W. T. (2010). The evolution of language, *New York: Cambridge University Press.*

Goodman, N. (1976). Languages of art. An approach to a theory of symbols *(2nd ed.), Indianapolis: Hackett Publishing Company Inc.*

Goodman, N. (1984). Of mind and other matters, Cambridge, MA: Harvard University Press.

Gratier, M. (2008). Liminal spaces and narratives of voice and body in infant vocal interchange (Commentary on Morioka), International journal for dialogical science, *3(1), 143–154.*

Gratier, M. & Trevarthen, C. (2007). Voice, vitality and meaning: On the shaping of the infant's utterances in willing engagement with culture. Comment on Bertau's "on the notion of voice", International journal for dialogical science, *2(1), 169–181.*

Gratier, M. & Trevarthen, C. (2008). Musical narrative and motives for culture in mother-infant vocal interaction. In C. Whitehead (ed.), The origin of consciousness in the social world *(pp. 122-158), Exeter, UK: Imprint Academic.*

Hanslick, E. (1989). Vom Musikalisch-Schönen. Ein Beitrag zur Revision der Ästhetik der Tonkunst *(16th ed.), Wiesbaden: Breitkopf und Härtel.*

Hurford, J. R. (2007). The origins of meaning, *Oxford: Oxford University Press.*

Huron, D. (2007). Sweet anticipation. Music and the psychology of expection, *Cambridge, MA: The MIT Press.*

Ingarden, R. (1968). Vom Erkennen des literarischen Kunstwerks, *Darmstadt: Wissenschaftliche Buchgesellschaft.*

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. Physics of life reviews, *8, 89–105.*

Kühl, O. (2007). Musical semantics, *Bern: Peter Lang.*

Langer, S. K. (1979). Philosophy in a new key. A study in the symbolism of reason, rite, and art *(3rd ed.), Cambridge, MA: Harvard University Press.*

Leman, M. (2008). Embodied music cognition and mediation technology. *Cambridge, MA: The MIT Press.*

Lerdahl, F. & Jackendoff, R. (1983). A generative theory of tonal music, *Cambridge, MA: The MIT Press.*

Levinson, S. C. (1983). Pragmatics. *Cambridge: Cambridge University Press.*

Levinson, S. C. (2013). Cross-cultural universals and communication structures. In M.A. Arbib (ed.), Language, music, and the brain: A mysterious relationship *(pp. 67-80), Cambridge, MA.: The MIT Press.*

Löbner, S. (1996). Bedeutung (meaning). In G. Strube (ed.). Wörterbuch der Kognitionswissenschaft *(pp. 55-57), Stuttgart: Klett-Cotta.*
Malloch, S. & Trevarthen, C. (eds.). (2009) Communicative musicality. Exploring the basis of human companionship, *Oxford: Oxford University Press.*
Margulis, E. H. (2014). On repeat. How music plays the mind. *New York: Oxford University Press.*
Martindale, C. (1999). Peak shift, prototypicality and aesthetic preference, Journal of consciousness studies, *6(6-7), 52–54.*
Meyer, L.B. (1956). Emotion and meaning in music, *Chicago: The University of Chicago Press.*
Morris, C. W. (1938). Foundations of the theory of signs, *Chicago: The University of Chicago Press.*
Morris, C. W. (1946). Signs, language, and behavior, *New York: George Braziller, INC.*
Newcomb, A. (1984a). Sound and feeling, *Critical Inquiry, 10(4), 614-643.*
Newcomb, A. (1984b). Once more "between absolute and program music": Schumann's second symphony, 19th-century music, *7(3), 233-250.*
Patel, A. D. (2008). Music, language, and the brain. *Oxford: Oxford University Press.*
Raffman, D. (1993). Language, music, and mind, *Cambridge, MA: The MIT Press.*
Ramachandran, V. S. & Hirstein, W. (1999). The science of art. A neurological theory of aesthetic Experience, Journal of consciousness studies, *6(6-7), 15–51.*
Ryan, M.-L. (2008). Narrative. In D. Herman, M. Jahn & M.-L. Ryan (eds.), Routledge encyclopedia of narrative theory *(pp. 344–348), London: Routledge.*
Schellekens, E. & Goldie, P. (eds.). (2011). The aesthetic mind. Philosophy and psychology. *Oxford: Oxford University Press.*
Sperber, D. & Wilson, D. (1986). Relevance. Communication and cognition, *Oxford: Basil Blackwell.*
Trevarthen, C. (2008). The musical art of infant conversation: Narrating in the time of sympathetic experience, without rational interpretation, before words, Musicae scientiae, special issue: Narrative in music and interaction, *15-46.*
Turner, M. (1996). The literary mind. The origins of thought and language, *Oxford and New York: Oxford University Press.*
Turner, M. (ed.). (2006). The artful mind. Cognitive science and the riddle of human creativity, *Oxford: Oxford University Press.*
Verhoef, T., Kirby, S. & Padden, C.(2011). Cultural emergence of combinatorial structure in an artificial whistled language, In L. Carlson, C. Hoelscher, & T. F. Shipley (eds.), Expanding the space of cognitive science: proceedings of the 33rd Annual Conference of the Cognitive Science Society: Boston, Massachusetts, July 20-23, 2011 *(pp. 483–488), Austin, TX: Cognitive Science Society.*
Whitehead, C. (2001). Social mirrors and shared experiential worlds, Journal of consciousness studies, *8(4), 3-36.*
Wolf, W. (2008). Music and narrative. In D. Herman, M. Jahn & M.-L. Ryan, (eds.), Routledge encyclopedia of narrative Theory *(pp. 324–329), London: Routledge.*
Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds, Current directions in psychological science, *8(1), 15–18.*
Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension, Journal of experimental psychology: learning, memory, and cognition, *21(2), 386–397.*

[11E] Listening and Attention

LOS ANGELES 14:00-16:00 Friday 8 Aug 2014

INSTRUMENT IDENTIFICATION IN CONCURRENT UNISON DYADS: THE EFFECT OF TIMBRE SALIENCY

Song Hui Chon[1], David Huron[1]
[1]School of Music, Ohio State University, USA
Time: 14:00

ABSTRACT

Saliency refers to the character of an object standing out, hence easily capturing someone's attention. This concept, when applied to music, can be interpreted in a way that a highly salient instrument sound will more likely draw a listener's attention than other concurrent instruments. That is, we might expect in turn that a highly salient instrument timbre will be more easily identified from other concurrent sounds, which will be also reflected by listeners' higher confidence ratings. In order to test these hypotheses, an experiment was carried out where twenty-seven listeners heard concurrent unison dyads of different instruments and were asked to identify one instrument present. Presented with a list of possible instruments, listeners identified one instrument they heard in a stimulus and rated their confidence that the chosen instrument was present. The results proved not to be consistent with the hypothesis that listeners are better at identifying those instruments that exhibit greater timbre saliency (Chon, 2013). Possible explanations for this failure are discussed. Still, a mild yet statistically significant correlation was observed between the timbre saliency values and average confidence ratings, which is consistent with one of our hypotheses that listeners would be more confident of identifying more salient instrument sounds.

1. INTRODUCTION

Timbre is widely regarded as the principal correlate of *sound source identity*. In the musical application of orchestration, timbre plays an important role in the phenomenon of *blend*. In the theory that motivates this research, both sound source identity and blend are conjectured to be influenced by a third concept, *timbre saliency*.

Timbre saliency refers to the attention-capturing quality of timbre (Chon & McAdams, 2012a). This definition can be interpreted in a way that a highly salient sound will more likely draw a listener's attention than other concurrent sounds. In the case of music, one might suppose that the degree of instrument saliency might have been an implicit factor in orchestration practices. Chon (2013) carried out a series of experiments to define timbre saliency and to examine its effect on perceived blend and voice recognition. Chon and McAdams (2012b) reported that the perceived degree of blend was negatively correlated with the timbre saliency values measured by Chon (2013), which was in agreement with the hypothesis that a salient timbre would not blend well.

There has been only a handful of research on timbre blend (Kendall & Carterette, 1993; Sandell, 1995; Tardieu & McAdams, 2011; Chon & McAdams, 2012b). Unlike most timbre research that focus on isolated sounds, blend studies focus on the interaction of concurrent sounds. In these studies, similar timbres were observed to blend better (Kendall & Carterette, 1993). The degree of perceived blend has been reported to be

negatively correlated with the identifiability of the underlying timbres (Kendall & Carterette, 1993). Lower spectral centroids (i.e., a darker sound) and a slower attack have been shown to be important in increasing the degree of blend (Sandell, 1995; Tardieu & McAdams, 2011; Chon & McAdams, 2012b).

In this study, the theory of timbre saliency will be used to examine the identification of an instrument in concurrent unison dyads. Specifically, we have two hypotheses:

H1. It is easier for listeners to identify timbres that are more salient. This ease of identification will be reflected in the average correct identification rate across listeners.

H2. Listeners are more confident of their choice of instrument identification with more salient instruments. This implies that there exists a positive correlation between the estimated timbre saliency scores of different timbres from Chon (2013) and the confidence ratings for listener identification of those sounds.

2. EXPERIMENT

2.1. Stimuli

The sounds used in the study were the same as those used in Chon and McAdams (2012a/b). These sounds include recorded samples of tones produced by 15 Western orchestral instruments from the Vienna Symphonic Library (VSL, 2011): Clarinet (CL), English Horn (EH), Flute (FL), French Horn (FH), Harp (HA), Harpsichord (HC), Marimba (MA), Oboe (OB), Piano (PF), Trombone (TN), Trumpet (TP), Tuba (TU), Tubular bells (TB), Violoncello (VC) and Vibraphone (VP). Each single-instrument tone was equalized in terms of pitch (C4, 261.6 Hz), loudness and effective duration. For this experiment, some stimuli consisted of single instrument tones. However, most stimuli consisted of combinations of two different timbres. In total, 120 stimuli were tested.

2.2. Participants

Three groups of participants were recruited based on self-identification: 10 professional musicians, 8 amateur musicians, and 9 non-musicians. Participants were undergraduate and graduate students and a faculty member from the Ohio State University community, including 20 females and 7 males. All of them reported normal hearing.

2.3. Procedure

Prior to the experiment, participants received an introduction to the experiment and gave a verbal consent. They then read the instructions on a printed sheet of paper. Participants were tested individually in an Industrial Acoustics Corporation sound attenuated room. Stimuli were heard through Sennheiser HD 280 Pro headphones at a comfortable listening level. The order of stimuli was randomized for each participant.

The entire experiment was implemented in MATLAB including the graphic user interface. The experiment consisted of three parts: training, preliminary test and main test. A “timbre training” display was provided for the training part with buttons whose labels identified each of the 15 individual timbres. After exploring these timbres at leisure, the participant proceeded to the preliminary test, where he/she was asked to identify each of the 15 timbres. Since the wind instruments were known in advance to be more easily confused with each other, the eight wind instrument sounds were tested twice, resulting in 23 trials. All trials were randomized in order.

Feedback was provided during the preliminary test. Participants were forced back to training and preliminary test until they had achieved a minimum of 65% correct identification to advance to the main experiment. This criterion was set a priori, prior to any data collection.

In the main test, participants were presented with either a single instrument timbre or concurrent unison dyads of two instrument timbres. The stimuli were presented in two blocks of 60 stimuli each, with a brief break available between the blocks. The task was to identify one of the instruments present in the stimulus and to provide the confidence rating for the indicated instrument, on a five-point scale from “not sure” to “very confident.”

In indicating the instrument heard in each stimulus, participants made a selection from a pull-down menu listing all 15 instruments. We were concerned that an alphabetic ordering of instruments might encourage participants to select instruments earlier in the alphabet, so we randomized the order of instruments on the menu for each stimulus. After selecting an instrument, the participant then selected the confidence level using one of five radio buttons. Stimuli were presented just once with no opportunity for the participant to replay a stimulus. This was done in order to avoid any unbalanced training effect from multiple exposures to more difficult (or more blended) stimuli.

3. RESULTS

Before starting on any analysis, we established an inclusion criterion. Recall that in the main experiment there were 15 single-instrument timbre trials, separate from 105 trials with concurrent unison dyads. If participants had difficulty identifying a single instrument in isolation, then results for identifying an instrument in paired stimuli would be much less reliable. Accordingly, a priori, we decided to include data for only the participants who scored better than 50% instrument recognition for single-instrument stimuli across the main experiment (a chance level would be 6.7%). As a result, 24 of 27 participants satisfied the inclusion criterion.

First, the responses for the single-instrument stimuli were examined. A confusion matrix has been computed from all of the single-instrument stimuli, including the results of 23 trials from the final preliminary test plus the 15 single-instrument stimuli dispersed through the main experiment (38 trials X 24 participants = 912 responses). Since each participant tested wind instruments twice and non-wind instruments once in the preliminary test, the values in Table 1 have been normalized accordingly. The timbre on each row is the instrument identified by the participants. On each column is the instrument timbre that was actually played. Table 1 is organized according to instrument families: the first four rows and columns are woodwinds; the next four are brass; the following three are percussions; the last four are string instruments, where HC and HA are plucked, PF is struck and VC is bowed. For each row or column, the largest value is observed on diagonal entries, indicating that the played instrument was identified correctly more often than not. However, the diagonal entries range from 0.417 corresponding to FH to 1 corresponding to VC, reflecting different degrees of ease in identifying 15 instruments.

Table 1: The confusion matrix from single-instrument timbre identification trials. On the row is the indicated instrument, whereas on the column is the actual played instrument. The instruments are listed according to the instrument families. (CL=clarinet, FL=flute, OB=oboe, EH=English horn,

TP=trumpet, FH=French horn, TN=trombone, TU=tuba, MA=marimba, VP=vibraphone, TB=tubular bells, HC=harpsichord, HA=harp, PF=piano, VC=violoncello)

	CL	FL	OB	EH	TP	FH	TN	TU	MA	VP	TB	HC	HA	PF	VC
CL	**0.85**	0.08	0.04	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FL	0.07	**0.83**	0.01	0.03	0.01	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OB	0.00	0.03	**0.57**	0.33	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
EH	0.01	0.00	0.21	**0.72**	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
TP	0.00	0.00	0.04	0.04	**0.69**	0.11	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FH	0.00	0.01	0.00	0.03	0.08	**0.42**	0.39	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TN	0.00	0.03	0.00	0.03	0.21	0.14	**0.58**	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TU	0.00	0.01	0.00	0.01	0.01	0.10	0.07	**0.74**	0.00	0.01	0.00	0.00	0.03	0.01	0.00
MA	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	**0.92**	0.06	0.00	0.00	0.00	0.00	0.00
VP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	**0.79**	0.02	0.04	0.00	0.00	0.00
TB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	**0.90**	0.02	0.00	0.00	0.00
HC	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.04	0.00	**0.92**	0.02	0.00	0.00
HA	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.02	0.00	0.00	**0.85**	0.04	0.00
PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	**0.98**	0.00
VC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**

The greatest confusion can be seen between FH-TN and between OB-EH, which happen to be the most similar timbre pairs according to the timbre dissimilarity study by Chon (2013). The confusion patterns exhibit a strong tendency to confuse instruments within a family. These results are comparable to the report by Giordano (2005) that listeners often confuse instrument tones produced by similar physical structure.

Since musicians are more likely to be familiar with different instruments, we might anticipate that the accuracy of instrument identification is associated with self-declared musicianship status. Accordingly, we calculated the per-group average identification rate of 15 timbres in isolation. There is a significant performance difference according to the group, $F(2,21)=10.672$, $p=.001$. A Bonferroni post-hoc test reveals that the difference between professional/amateur versus non-musician was found to be significant ($p=.001$ and $p=.020$, respectively). There was no significant difference between self-declared professional and amateur musicians ($p=.963$). This result suggests that self-declared musicians in general performed better in the given task, which is in agreement with participants' post-experiment interviews.

Addressing the main experimental hypothesis, we next consider whether timbre saliency influences instrument identification for isolated tones. Using the timbre saliency values from Chon (2013), we calculated the correlation between saliency values and average percent correct identification rates across all participants. Contrary to the experimental hypothesis, no statistically significant relationship is evident, $r(13)=.39$, $p=.1562$ for timbres in isolation; $r(13)=.27$, $p=.3372$ for timbres presented with others.

Our main experimental hypothesis predicts that timbre saliency would contribute to instrument identification. Conceptually, we might suppose that a number of factors influence the ability of a listener to identify the presence of an instrument in a unison dyad. First, we might expect that the more salient an instrument, the more likely it would be identified. However, the ability to identify the more salient instrument might be influenced by the saliency of the other instrument in the pair. If the two instruments have quite similar saliency values, then the competition is apt to be high. Conversely, if there is a large difference in the saliency of the two instruments, one might predict that the more salient instrument will benefit, and therefore be easier to identify. That is, the *difference* in saliency might be expected to influence listener judgments.

In order to test our main hypothesis, we constructed an a priori formal model that includes these proposed effects and interactions. Specifically, our model includes four factors, which can be interpreted as predictor variables in a multiple regression analysis:

$$p(X \mid X+Y) = c_1 p(X) + c_2 S_X + c_3 S_{diff}$$

where $p(X|X+Y)$ is the probability of identifying instrument X when presented concurrently with instrument Y, $p(X)$ is the probability of identifying the target instrument X in isolation, S_X is the saliency value of the target instrument X, S_{diff} is the absolute difference in saliency with respect to the other instrument. S_X was obtained by adding 0.5 to the saliency values from Chon (2013), because the original values ranged from -0.5 to +0.5. By shifting all values to the [0, 1] range, S_X now has the same range with other variables in the equation above.

Using the ENTER method of multiple regression, the adjusted R-squared value was determined 0.322, with the regression model being statistically significant to the accounted variance, $F(3,206)=34.067$, $p<.000001$. However, a close examination revealed that $p(X)$ was the only significant factor in predicting the correct identification of an instrument timbre in dyads, $beta=.575$, $p<.000001$. Neither timbre saliency values nor saliency differences had a significant impact.

One of the hypotheses was that listeners would be more confident of their identification with more salient timbres. In order to address this hypothesis, we calculated the correlation between the saliency values from Chon (2013) and the average confidence ratings across all participants. The result was a mild but statistically significant correlation, $r(208)=.2405$, $p<.001$, which is in agreement with our hypothesis.

4. DISCUSSION

In this paper, we examined the identification of an instrument sound in concurrent unison dyads. As a salient timbre is defined to be the one that captures listeners' attention easily and tends not to blend well with concurrent sounds (Chon & McAdams, 2012a/b), we can then logically expect that a salient timbre will be easily identified. In that sense, this experiment can be considered as a comparison with timbre saliency previously measured by Chon and McAdams (2012a).

In the earlier study, Chon and McAdams (2012a) used a tapping technique to a perceptually isochronous ABAB sequence where A and B were two different timbres with their pitch, loudness and effective durations equalized. The result was a one-dimensional timbre saliency scale where the distance between a pair of timbres would be proportional to the saliency difference between them. This scale exhibited a negative correlation with the perceived blend of concurrent unison dyads in a later study (Chon & McAdams 2012b).

As a highly salient timbre was reported to show little blend with a concurrent sound (Chon & McAdams 2012b), we hypothesized that a highly salient sound will also be easy to identify. Hence, the identification task can be considered as a measure of timbre saliency. Consequently, we expected result that would be confirmative of earlier measured saliency values by Chon (2013). Contrary to our expectation, the measured identification result failed to show a statistically significant correlation with the previous timbre saliency values or the saliency differences. The only significant factor in predicting the correct identification of a timbre in concurrent dyads turned out to be the correct identification score in isolation. In other words, a sound that was easy to identify in isolation was also easy to identify when presented with another concurrent sound, and vice versa. This may sound trivial, but to our knowledge this is the first study to notice this correlational relationship.

Our data were consistent with the other hypothesis, namely, that listeners would be more confident of their identification of a more salient timbre. However, if we grant that the measured identification score is another measure of saliency, the correlation of measured saliency (i.e., the average identification score) and the confidence rating becomes very strong, $r(208)=.9899$, $p<10^{-10}$. This much higher correlation seems to make sense, as this is a result of two dependent variables obtained from the same experiment, whereas the timbre saliency values were collected from another experiment. This difference in correlation magnitudes might suggest a context effect. It is not known yet how a measure of saliency in one context can be applied to another context. Nevertheless, in summary, regardless of which measure of saliency to use, there is likely a statistically significant relationship between saliency and ease of identification.

In post-experiment interviews many participants mentioned a common set of timbres that were easy to identify and a different set of timbres that were difficult to identify. Easily identifiable instruments included the VC (the only bowed string instrument that had some uncharacteristic nasal quality due to a high pitch on a cello), PF, HC and "the bells" (MA, VP, TB). Participants were unanimous in reporting the brass instruments as the most difficult to identify. FH-TN and OB-EH pairs were the most challenging to distinguish, even by musicians who were trained on those instruments.

In comparison with the experiment by Chon and McAdams (2012a) to measure saliency, our participants reported different instruments to be more easily identifiable. Most noticeable is the case of VC. VC, being the only bowed string instrument, was identified more easily than any other instrument in this experiment, whereas it was one of the least preferred ones in the tapping task. This discrepancy might come from the multidimensionality of timbre, in the sense that different contexts bring out different characters of timbre, which participants unconsciously latch on to.

The correct identification score in dyads also showed a significant and very strong correlation with the confidence ratings, which additionally showed a mild yet statistically significant correlation with the timbre saliency values. If we consider the identification scores as another measure of saliency, both significant correlations are in agreement with our hypothesis that listeners would be more confident of identifying more salient timbres.

5. REFERENCES

Chon, S.H. (2013). Timbre Saliency, The Attention-Capturing Quality of Timbre. Unpublished Ph.D. Dissertation. McGill University, Canada.

Chon, S.H. & McAdams, S. (2012a). Investigation of Timbre Saliency, the Attention-Capturing Quality of Timbre, Journal of Acoustical Society of America, 131*(4), 3433 – 3433.*

Chon, S.H. & McAdams, S. (2012b). Exploring Blending as a Function of Timbre Saliency, In Proceedings of the 12th International Conference of Music Perception and Cognition (ICMPC), *Thessaloniki, Greece (pp. 222 – 223).*

Giordano, B.L. (2005). Sound source perception in impact sounds. Unpublished Ph.D. Dissertation. University of Padova, Italy.

Kendall, R., & Carterette, E. (1993). Identification and blend of timbres as a basis for orchestration. Contemporary Music Review, 9, *51–67.*

Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. Music Perception, 13, *209–246.*

Tardieu, D., & McAdams, S. (2012). Perception of dyads of impulsive and sustained sounds. Music Perception, 30*(2), 117–128.*

Vienna Symphonic Library GmbH. (2011). Vienna symphonic library. Available from http://vsl.co.at.

PERCEPTUAL SALIENCE OF CADENTIAL AND NON-CADENTIAL GLIDES IN SEMITONE SEQUENCES

Jiaxi LIU[1]

[1]*University of Cambridge, United Kingdom*

Time: 14:30

A previous corpus study on performed violin glides between semitone sequences suggested that pitch glides of 28% and 11% of the leading tone were respectively associated with cadential and non-cadential function. Two perceptual experiments were carried out to assess their discriminability and preference. Semitone stimuli were generated using sawtooth waveforms with a linear pitch glide between the notes taking up 11% or 28% of the leading tone duration. Stimuli were additionally varied in tempo (slow, moderate, and fast) and final note duration ratio (1:1 and 1:2). In the discrimination task, 30 participants identified the pair that contained different glide durations in a 4IAX presentation. All participants discriminated the two glide durations above chance, with musicians performing better for the 28% than the 11% glide (p=0.002). This

suggests that musicians used their training and experience to notice the salient cadential structural difference from the norm. In the preference task, two notes preceded the original stimulus to make up a new 4-note sequence that allowed a cadential ('sol-la-ti-do') or a non-cadential ('do-re-mi-fa') hearing of the phrase-final semitone. Glide duration, tempo and final note durations were varied as before. In an AX presentation, the same 30 participants indicated their preference for one or the other glide duration. All participants preferred the 11% glide duration regardless of context, although their preference for the 28% glide increased for the slow tempo ($p=0.0037$) and a longer final note. These results are consistent with slower and longer final note characteristics of musical cadences, for which the 28% glide would be appropriate. Musicians were also more consistent in their choice than non-musicians. Together, these findings show that listeners do distinguish between and show preference for pitch glide durations. That musicians performed better in both tasks suggests that explicit musical training may influence both discriminability and preference for certain musical structural marker associations such as pitch glides. In order to minimise the noise in the data potentially caused by the artificial stimuli used, future experimentation with more ecologically-valid live violin recordings would be beneficial.

THE EFFECT OF MUSICAL TRAINING ON AUDITORY GROUPING

Sarah Sauve[1], *Lauren Stewart*[1], *Marcus Pearce*[2]
[1]*Goldsmiths, University of London, United Kingdom,* [2]*Queen Mary, University of London, United Kingdom*
Time: 15:00

ABSTRACT

Auditory scene analysis is the process by which everyday sound environments are understood; auditory streaming is highly relevant to this process, particularly with respect to timbre, separating incoming sounds by source to identify danger or opportunity in one's surroundings. Of the various parameters that influence auditory streaming (Bregman, 1990), we focus on musical training with respect to timbre and attention. An increase in performance in many auditory tasks is seen in musicians, including streaming (Zendel & Alain, 2009), presumably a result of training and brain plasticity. This experiment corroborates this observed effect of musical training and clearly demonstrates an influence of attention on streaming while suggesting further effects of training on specific instruments. Both non-musicians and specific instrumentalists took part in a simple ABA-paradigm where timbre was manipulated. The manipulation of attention formed two boundaries, identified as the fission boundary and the temporal coherence boundary, significantly different between musicians and non-musicians and additionally affected by specific timbres.

1. INTRODUCTION

Auditory streaming has been investigated in the context of numerous sound attributes such as pitch (van Noorden, 1975), location (Jones & Macken, 1995), periodicity (Vliegen, Moore, & Oxenham, 1999) and timbre (Iverson, 1995) among others. The role of musical training has also been extensively studied in the context of auditory skills, including auditory streaming (Zendel & Alain, 2009). As a result of training, musicians are more sensitive to changes in auditory stimuli based on pitch, time and loudness for example (Marozeau, Innes-Brown, & Blamey, 2013), with discrimination thresholds being lower in musicians than in non-musicians. One problem with treating musicians as one single category is that fine differences between instrumentalists may be missed. Pantev and colleagues (Pantev, Roberts, Schulz, Engelien, & Ross, 2001) found that certain instrumentalists were more sensitive to the timbre of their own instrument than to others, measured by auditory evoked fields. However, this has not yet been observed in auditory streaming, where an effect would be seen by a change in streaming threshold; presumably, it would take less change to recognize two separate auditory objects when one's own instrumental timbre is one of these objects.

Over the last few decades, the emerging concept of auditory streaming is that of a two-staged process: the first being pre-attentive, peripheral and following Gestalt principles (Bregman, 1990) and the second post-attentive, cortical and involving top-down, heuristic and schematic mechanisms (Alain, Arnott, & Picton, 2001). Though there is still ongoing debate about the role of attention in streaming, specifically whether steaming is a pre- or post-attentive phenomenon, it is clear that attention influences perception (van Noorden, 1975). With accumulating evidence for a two-staged process (Snyder & Alain, 2007; Winkler, Takegata, & Sussman, 2005), a better question now is: what is the influence of attention at each stage?

Following the rules of Gestalt psychology, auditory streaming displays multi-stability. Within certain criteria, listeners can hear either coherence or segregation, based on top-down processes such as training, attention or expectation (Bendixen, Denham, Gyimesi, & Winkler, 2010). van Noorden (1975) defined two boundaries within which bi-stability was possible by manipulating attentional focus. These boundaries are the fission boundary, the difference in pitch below which segregation into two auditory objects, or streams, is not possible, and the temporal coherence boundary, the difference in pitch above which coherence into one auditory object, or stream, is no longer possible.

This research has three objectives. The first is to reveal the fission and temporal coherence boundaries in streaming by timbre through a manipulation of attentional focus. The second is to test whether these boundaries are influenced by musical training, with boundaries hypothesized to be lower for musicians. The third is to find behavioural evidence for an increased sensitivity in streaming the instrument(s) which a musician plays, expected to be shown by lower boundaries for piano sounds in pianists for example.

2. METHOD

2.1. Participants

Participants were 21 university students, 12 females and 9 males, average age 26.19 (SD = 8.47). 15 of these participants were musicians and 6 were non-musicians, determined by Gold-MSI scores. Of these musicians, four named the piano as their primary instrument while three named the violin. The remaining eight named a range of other instruments, including voice.

2.2. Stimuli

All six timbral sounds (piano, violin, trumpet, trombone, clarinet, bassoon) were chosen from the MUMS library (Opolko & Wapnick, 2006) to be of equal pitch (A440) and were adjusted to equal perceptual length of 100ms and equal loudness, based on the softest sound. A 10ms fade out was applied to each timbral sound, all edits using Audacity and exported as a wave file.

The 'standard' sequence was presented at a rate of onset of 220ms and did not change within a trial. The 'target' sequence was presented at a rate of onset of 440ms, beginning 110ms after the 'standard' sequence to create the well-known galloping ABA pattern (van Noorden, 1975). The target sequence was a 30s cross-fade between the standard timbre and the target timbre. The target sequence cross-faded from standard to

target timbre in the ascending condition, creating a galloping to even rhythm change, and from target to standard timbre in the descending condition, creating an even to galloping rhythm change (Figure 1). Each trial ended when the participant indicated a change in perception.

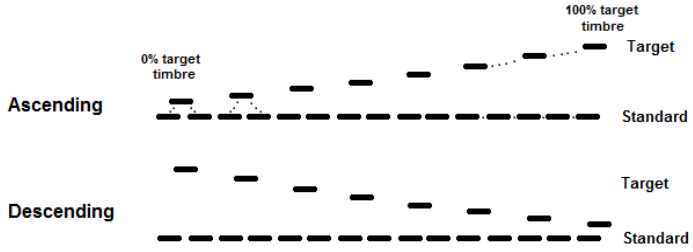


Figure 1: Illustrated stimulus for an ascending trial and a descending trial, with time represented horizontally and timbre represented vertically.

2.3. Procedure

he experiment was coded and run in Max/MSP, with output heard through headphones and input taken from mouse clicks. Participants were first presented with a practice patch with instructions and an opportunity to listen to each timbre and rhythm separately. Up to twelve practice trials were included in the patch and questions were welcomed.

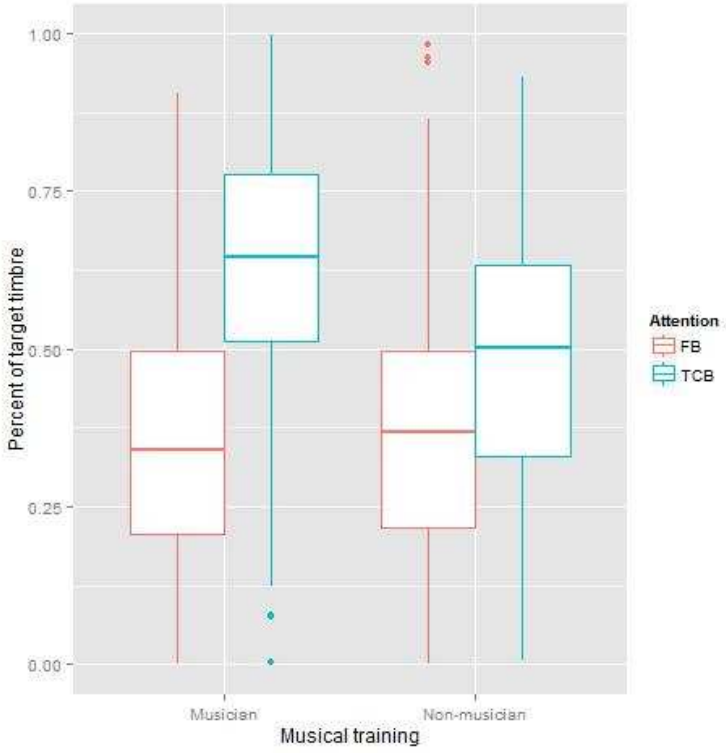


Figure 2: Percent of target timbre as a function of musical training and attention, with a clear division between FB and TCB. The error bars show the first and third quartiles.

For each trial, participants indicated by clicking a button on the screen at which point the galloping sequence becomes perceived as two separate streams of standard and target tones, or the opposite for decreasing presentation. This point was recorded as percent of target timbre present in the cross-fade at that time. Each trial lasted a maximum of 30s, at which point the trial ended itself and recorded a value of '-1' for that trial, indicating that the participant had never reached a change in perception. The manipulation of attention was as follows: in two blocks, the instructions were to indicate a change in rhythm as soon as it was perceived and for the other two blocks the instructions were to hold on to the original rhythm as long as possible. Each pair of blocks contained an ascending and descending block, for a total of four blocks. For every block, every timbre modulated to every other timbre once for a total of 30 trials (6 timbres each modulating to the 5 other timbres), each separated by 4s. Participants were assigned to different orders (four possibilities) in rotation to prevent order effects.

Additionally, once all blocks were completed, participants filled out the musical training sub-scale of the Goldsmiths Musical Sophistication Index.

3. RESULTS

Two non-musicians' data were removed from analysis as the task was clearly not understood, leaving 4 non-musicians. Also, trials where participants did not reach a change in perception were removed from the main data and analyzed separately. The trumpet and trombone timbres were grouped together as 'brass' and the clarinet and bassoon timbres were grouped together as 'woodwinds' for the analysis.

The dependent variable was the percentage of target timbre present in the cross-fade of the target sequence at the time the participant indicated a change in perception, expressed in decimal value. A between-subjects ANOVA for order effects (4 levels) was non-significant ($p > .05$), as expected. An independent-samples t-test for direction was non-significant ($p > .05$), also as expected. A mixed ANOVA with musical training (2 levels) as a between-subjects variable and attentional focus (2 levels) as a within-subjects variable yielded a main effect of musical training, F (1, 17) = 5.53, p = .03, attentional focus, F (1, 17) = 52.22, $p < .01$, and a significant interaction, F (1, 17) = 4.87, p = .04 (Figure 2), with attentional focus having stronger effect on musicians. As attentional focus was significant, further analysis will be conducted for the fission boundary (FB) and temporal coherence boundary (TCB) as separate phenomena. For all trials where piano was the standard timbre, a between-subjects ANOVA for musical training (4 levels) was conducted with non-significant results for both the FB and the TCB (both $p > .05$). For all trials where piano was the target timbre, the same ANOVA was conducted with a significant effect for the TCB, F (3, 15) = 6.18, $p < .01$, but not for the FB ($p > .05$); however, both tests failed Levene's test for homogeneity of variance. Therefore, a post-hoc pairwise t-test with Bonferroni correction was conducted revealing a significant difference between pianists and violinists for the FB, $p < .01$ (Figure 3a), and significant differences between musicians and non-musicians for the TCB.

Similarly, for all trials where violin was both the standard and the target timbre, a between-subjects ANOVA for musical training (4 levels) was conducted for the FB and the TCB. In all cases, Levene's test of homogeneity of variance failed; therefore pairwise comparisons with Bonferroni correction were applied. There was a significant difference between pianists and violinists at the FB when violin was the standard, p = .02 (Figure 3b). There was a significant difference between all types of musician and non-musicians at the TCB when violin was the standard and when violin was the target, $p < .01$. There was a significant difference between pianists-violinists and pianists-non-musicians at the FB when violin was the target timbre, p = .05 and $p < .01$ respectively (Figure 3c).

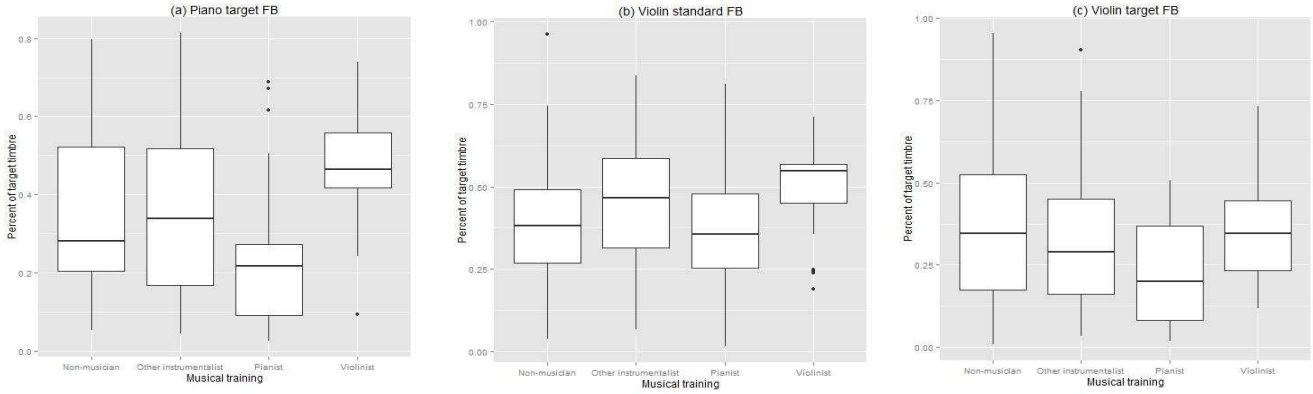


Figure 3: There were significant differences between pianists and violinists when piano was the target timbre (a), and violin was the standard (b) and target (c) timbre, all for the fission boundary. Error bars represent first and third quartiles.

Performance was also analyzed by participant group. An ANOVA for standard timbre (4 levels) was conducted on all pianists, with no significant effect of standard timbre on FB, after sphericity correction and no significant effect of standard timbre on TCB. A pairwise t-test with Bonferroni correction revealed a significant difference on the FB between piano/violin, and woodwinds/brass pairs. An ANOVA for target timbre (4 levels) was also conducted, revealing a significant effect of target timbre on the FB, after sphericity correction, F (3, 9) = 6.95, p = .01. A similar ANOVA for target timbre on the TCB revealed a significant effect of timbre, F (3, 9) = 5.57, p = .01, however pairwise comparisons were not significant.

The same analysis was performed for violinists, where an ANOVA for standard timbre on the FB, as well as on the TCB was non-significant. An ANOVA for target timbre on the FB revealed a main effect of timbre, F (3, 6) = 6.71, p = .02, no significant pairwise comparisons. There was no significant effect on the TCB.

Thresholds for all possible pairs of timbres were compared to see if performance was better for any particular timbral pairing; results were non-significant for both the FB and the TCB.

The trials where participants never reached a change in perception were analyzed by counting the number of instances by variable. First, the count was much higher for the TCB than the FB, as expected, $\chi 2$ (1) = 17.96, $p < .01$. More specifically, a lack of change only happened in the direction of the TCB for pianists. For violinists, a lack of change only happened once in the direction of the FB, indicating that other- and non-musicians carry all the cases of overshooting in the direction of the FB. The number of instances was not significantly different across different standard or target timbres but did happen for every possible timbral pair. The occurrence of timbral pairs was not evenly spread, $\chi 2$ (14) = 25.43, p = .03, with perception not changing most often for piano-violin, piano-clarinet, piano-bassoon and violin-trombone pairs of timbres.

4. DISCUSSION

The results have demonstrated that shifting the focus of attention has a clear effect on perception. Instructions to indicate the first hint of a change in rhythm or to hold on to the original rhythm as long as possible translate into focusing on the galloping rhythm (integration), which defines the temporal coherence boundary, or the even rhythm (segregation), which defines the fission boundary. This contradicts Carlyon and colleagues (Carlyon, Cusack, Foxton, & Robertson, 2001), who questioned the existence of the fission boundary while finding a clear temporal coherence boundary in a paper where streaming was measured outside the focus of attention. Perhaps if attention is focused away from the auditory scene, there are no separate boundaries but only streaming or lack thereof. This interpretation suggests that streaming can and does

occur outside the focus of attention but is treated with finer detail inside the focus of attention, reflecting differing function between the proposed pre- and post-attentive stages of auditory scene analysis (Snyder & Alain, 2007).

The results also show a clear difference in performance between musicians and non-musicians, though the observed effect is not as predicted. While discrimination boundaries were expected to be lower in musicians, the fission boundary is equal and the temporal coherence boundary significantly higher. This equates to a larger range of bi-stability, perhaps due to top-down processes having a stronger influence in musicians. Expectation can play a large role in influencing perception (Summerfield & Egner, 2009), especially with the use of directive instructions (van Noorden, 1975), and this may be stronger for musical concepts in those with musical training. This is supported by the fact that a lack of change in rhythm only occurred in the direction of the FB for non-musicians and other musicians, all having a lower Gold-MSI musical training score than the pianist and violinist participants. One theory of auditory streaming states that coherence is the default percept, with stream segregation only occurring with accumulated supporting evidence (Bregman, 1990); however, if this was the case the descending blocks of this design would not have resulted in any change in rhythm in any trial, demonstrating the strength of expectation. Instead, it seems likely that the expectation of hearing two streams and an even rhythm allows immediate perception of segregation. Similarly, musicians have a firmer schema for rhythm and timbre and could hold coherence over larger differences when instructed to. To implicate expectation and its influence with certainty, it would be interesting to compare responses to a simple one/two stream judgment task between the current set of instructions and the absence of leading instructions.

The data do not clearly demonstrate an advantage of musical training on a specific instrument to detecting specific timbre. Furthermore, the influence of timbre is not as apparent on the TCB as it is on the FB. This may be caused by a ceiling effect on the TCB in pianists and violinists, as we know that these particular instrumentalists did not necessarily reach a change in perception by the end of the trial, indicating an underestimation of the true location of the TCB in these instrumentalists. This makes it difficult to identify finer differences between musicians and poses a methodological challenge, as the two sequences are 100% separate timbres by the end of each trial and therefore cannot be made any more different. Once again, top-down processes could account for this, where musicians could be capable of holding together two different lines more steadily and reliably than non-musicians, as a result of training. Results did show that pianists generally performed at a lower threshold for piano sounds, but so did violinists and other musicians. Perhaps piano is common to all musicians, and details will come out more strongly when comparing performance on the pairs of brass and woodwind instruments against brass and woodwind players, as these pairs of instruments are both closer in timbral space (McAdams, Winsberg, Donnadieu, & De Soete, 1995) than piano and violin. No particular timbral pair statistically stood out as easier to discriminate than another; however, in informal discussion following the experiment, several participants reported certain timbral pairings to be much easier to segregate than others (for example piano-violin as opposed to trumpet-trombone).

In conclusion, this research has demonstrated that when attention is focused on the auditory stimulus, the streaming process can be influenced by manipulating the focus of attention. It has shown that musicians and non-musicians respond to this manipulation differently, with musicians having a wider range of acceptance for bi-stability. Recruiting brass and woodwind players is the next step in investigating the influence of timbre-specific training. Finally, it has also made clear that streaming cannot be discussed separately from attention and expectation, whose roles and influence remain to be specified.

5. REFERENCES

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. Journal of Experimental Psychology-Human Perception and Performance, 27*(5), 1072-1089. doi: 10.1037//0096-1523.27.5.1072*

Bendixen, A., Denham, S. L., Gyimesi, K., & Winkler, I. (2010). Regular patterns stabilize auditory streams. Journal of the Acoustical Society of America, 128*(6), 3658-3666. doi: 10.1121/1.3500695*

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. *Cambridge: MIT Press.*

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. Journal of Experimental Psychology-Human Perception and Performance, 27*(1), 115-127. doi: 10.1037//0096-1523.27.1.115*

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 751-763. doi: 10.1037/0096-1523.21.4.751*

Jones, D. M., & Macken, W. J. (1995). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. Memory & Cognition, 23*(2), 192-200. doi: 10.3758/BF03197221*

Marozeau, J., Innes-Brown, H., & Blamey, P. J. (2013). The effect of timbre and loudness on melody segregation. Music Perception, 30*(3), 259-274. doi: 10.1525/mP.2012.30.3.259*

McAdams, S., Winsberg, S., Donnadieu, S., & De Soete, G. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research/Psychologische Forschung, 58*(3), 177-192. doi: 10.1007/BF00419633*

Opolko, F., & Wapnick, J. (2006). The McGill University Master Samples Collection on DVD (3 DVDs). Quebec, Canada: McGill University.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport, 12*(1), 169-174. doi: 10.1097/00001756-200101220-00041*

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. Psychological Bulletin, 133*(5), 780-799. doi: 10.1037/0033-2909.133.5.780*

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. Trends in Cognitive Sciences, 13*(9), 403-409. doi: 10.1016/j.tics.2009.06.003*

van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences. Unpublished doctoral dissertation. Technische Hogeschool Eindhoven. Eindhoven, The Netherlands.

Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. Journal of the Acoustical Society of America, 106*(2), 938-945. doi: 10.1121/1.427140*

Winkler, I., Takegata, R., & Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Cognitive Brain Research, 25*(1), 291-299. doi: 10.1016/j.cogbrainres.2005.06.005*

Zendel, B. R., & Alain, C. (2009). Concurrent Sound Segregation Is Enhanced in Musicians. Journal of Cognitive Neuroscience, 21*(8), 1488-1498. doi: 10.1162/jocn.2009.21140*

DIFFERENCE OF OPTIMUM LISTENING LEVEL OF MUSIC AND PERCEIVED LOUDNESS OF SOUND BETWEEN MEN AND WOMEN

Mariko Hamamura[1], *Shin-Ichiro Iwamiya*[1]
[1]*Faculty of Design, Kyushu University, Japan*
Time: 15:30

ABSTRACT

To determine the factor affecting the difference of optimum listening level of music between men and women, the measurement experiment of optimum listening level and the rating experiment of perceived loudness of broadband and various center frequencies' narrowband noise were conducted.

The optimum listening levels of music of male participants were higher than those of female participants. The rating experiments on loudness showed that the female participants rated higher loudness scores than the male participants for broadband and narrowband noises of various center frequencies of the same sound pressure levels. This tendency suggested that the optimum listening level of music was affected by the loudness perception of sound. The difference of optimum listening level of music between men and women was due to the difference of perceived loudness of sound between men and women.

1. INTRODUCTION

Previous studies indicated the difference of optimum listening level of music between men and women (Barrett and Hodgets, 1995; Fligor and Ives, 2006; Hodgetts et al., 2007; Hamamura and Iwamiya, 2013). All of them reported that the optimum listening level of men was higher than that of women. However, the factors affecting the difference of optimum listening level between men and women were not clarified yet. Furthermore, the experimental conditions of the previous studies were not systematically enough to discuss the difference of optimum listening level between men and women.

In the present study, the optimum listening level of music was measured by systematic psychoacoustical experiment to confirm the difference of the optimum listening level of music between men and women. Furthermore, the rating experiment of perceived loudness of broadband and various center frequencies' narrowband noise was conducted to clarify the factor affecting the difference of the optimum listening level between men and women. The difference of loudness perception between men and women was supposed to be such a factor.

2. MEASUREMENT EXPERIMENT OF OPTIMUM LISTEING LEVEL OF MUSIC

To discuss the difference of the optimum listening level of music between men and women, the measurement experiment of the optimum listening level of music was conducted by using a method of adjustment.

2.1. Experimental condition and procedure

The sound stimuli were 4 music experts (approximately 90 s in duration). The information of each stimulus is shown in Table 1. The participants listened to the stimuli via headphones and adjusted the sound volume of headphone amplifier to the optimum listening level for enjoying each tune. In the previous studies, the optimum listening level was measured only once for each stimulus (Barrett and Hodgets, 1995; Fligor and Ives, 2006; Hodgetts et al., 2007; Hamamura and Iwamiya, 2013). However, if we adjusted the optimum listening level several times, the optimum listening level might be different even if the tune was the same. In the present study, the difference among the adjustment levels of each participant's trials for each stimulus was compared with the difference of the average adjustment level between the male and female participants. To discuss such differences, the optimum listening level for each stimulus was measured 5 times. The experiment was consisted of 5 sessions and 4 sound stimuli were reproduced randomly in each session. To avoid the case that the optimum listening level adjusted in previous sessions influence to the adjustment of current optimum listening level, the headphone amplifier and participants' hand were covered by a box.

The participants were 14 students of Kyushu University (7 males and 7 females), aged 21 to 31. All of them had normal hearing but difference of hearing thresholds at some frequencies between the male and female participants were statistically significant by Man Whitney's U test (250 Hz : U =50.5, $p < 0.05$; 500 Hz : U = 39.5, $p < 0.01$; 6 kHz : U = 52.5, $p < 0.05$). In these conditions, the hearing thresholds of the male participants were lower than those of the female participants.

Table 1: The information of music used as the sound stimuli for measurement of the optimum listening level.

No.	Title and artist	Genre
1	Numb, Linkin Park	American Rock
2	Sing, Sing, Sing, Benny Goodman	Jazz (Big band)
3	Linda Linda, THE BLUE HEARTS	Japanese Punk
4	Rock and Roll, Led Zeppelin	British Rock

2.2. Results

In this paper, the difference of the optimum listening levels within a session was called as "intra-individual variation". The intra-individual variation was defined as the difference between the maximum and minimum optimum listening level in each session. The intra-individuals variations and the average values of optimum listening level of 5 sessions for each participant for each stimulus are shown in Table 2. Table 2 also shows the average optimum listening levels for the male participants and those for the female participants for each stimulus. In Table 2, the

male participants are shown as M and the female participants are shown as F. The subsequent number of each alphabet means participants' number. For example, M1 means the male participant 1, F2 means the female participant 2.

As shown in Table 2, the male participants adjusted the optimum listening level higher than that of the female participants for all stimuli. This tendency is the same to the previous studies (Barrett and Hodgets, 1995; Fligor and Ives, 2006; Hodgetts et al., 2007; Hamamura and Iwamiya, 2013). The differences of the optimum listening levels between the male and female participants were larger than intra-individual variations expect M7 for tune 1 and M4 for tune 3.

In Table 3, 95 % confidence intervals for intra-individual variations are shown. The upper limits of 95 % confidence intervals for intra-individual variation are smaller than the differences of optimum listening level between the male and female participants for all the stimuli. These results showed that intra-individual variation of the optimum listening level caused by adjusting it several times was smaller than the difference of optimum listening level between the male and female participants.

Table 2: The average of optimum listening level and intra-individual variation for each stimulus [dB].

	Tune 1		Tune 2		Tune 3		Tune 4	
	Average of optimum listening level	Intra-individual variation	Average of optimum listening level	Intra-individual variation	Average of optimum listening level	Intra-individual variation	Average of optimum listening level	Intra-individual variation
M1	61.6	2.9	60.4	5.3	58.8	4.5	62.0	4.0
M2	62.1	3.5	74.0	0.1	66.0	7.0	72.5	6.3
M3	65.6	4.7	64.2	3.7	66.4	8.8	73.9	3.3
M4	76.5	9.0	80.8	9.8	75.0	12.6	80.5	9.7
M5	60.4	4.0	62.5	5.6	59.9	1.9	60.8	3.5
M6	63.2	6.5	61.8	4.3	62.1	4.6	64.9	3.7
M7	74.4	14.3	77.9	11.0	78.3	6.8	77.3	7.0
F1	67.7	3.6	69.5	3.6	65.6	2.8	67.4	2.3
F2	49.1	9.0	50.3	11.2	48.4	4.5	48.2	8.1
F3	51.2	4.0	52.6	4.4	50.3	4.7	51.0	5.2
F4	60.1	1.7	62.2	4.5	59.1	4.0	62.5	4.5
F5	45.6	7.2	48.3	10.4	46.0	3.9	49.3	9.7
F6	59.1	2.4	60.0	2.9	60.1	6.4	59.4	1.5
F7	58.4	4.9	58.7	5.5	55.2	4.9	58.6	5.3
Average of male	66.3		68.8		66.6		70.3	
Average of female	55.9		57.4		54.9		56.6	
Difference between male and female	10.4		11.4		11.7		13.7	

Table 3: 95 % confidence intervals for intra-individual variation in each stimulus [dB].

	Tune 1	Tune 2	Tune 3	Tune 4
Upper limit	7.4	7.8	7.1	6.8
Lower limit	4.0	3.9	4.0	3.8

The two-way ANOVA was applied to the optimum listening level of music. As a result, the effect of gender was statistically significant ($F(1,272) = 177.88$, $p < 0.01$) and the male participants adjusted the optimum listening level higher than the female participants. The effect of tune and the interaction between gender and tune were not statistically significant.

The experimental results confirmed the existence of the difference of optimum listening level between men and women. This difference was not observed coincidentally.

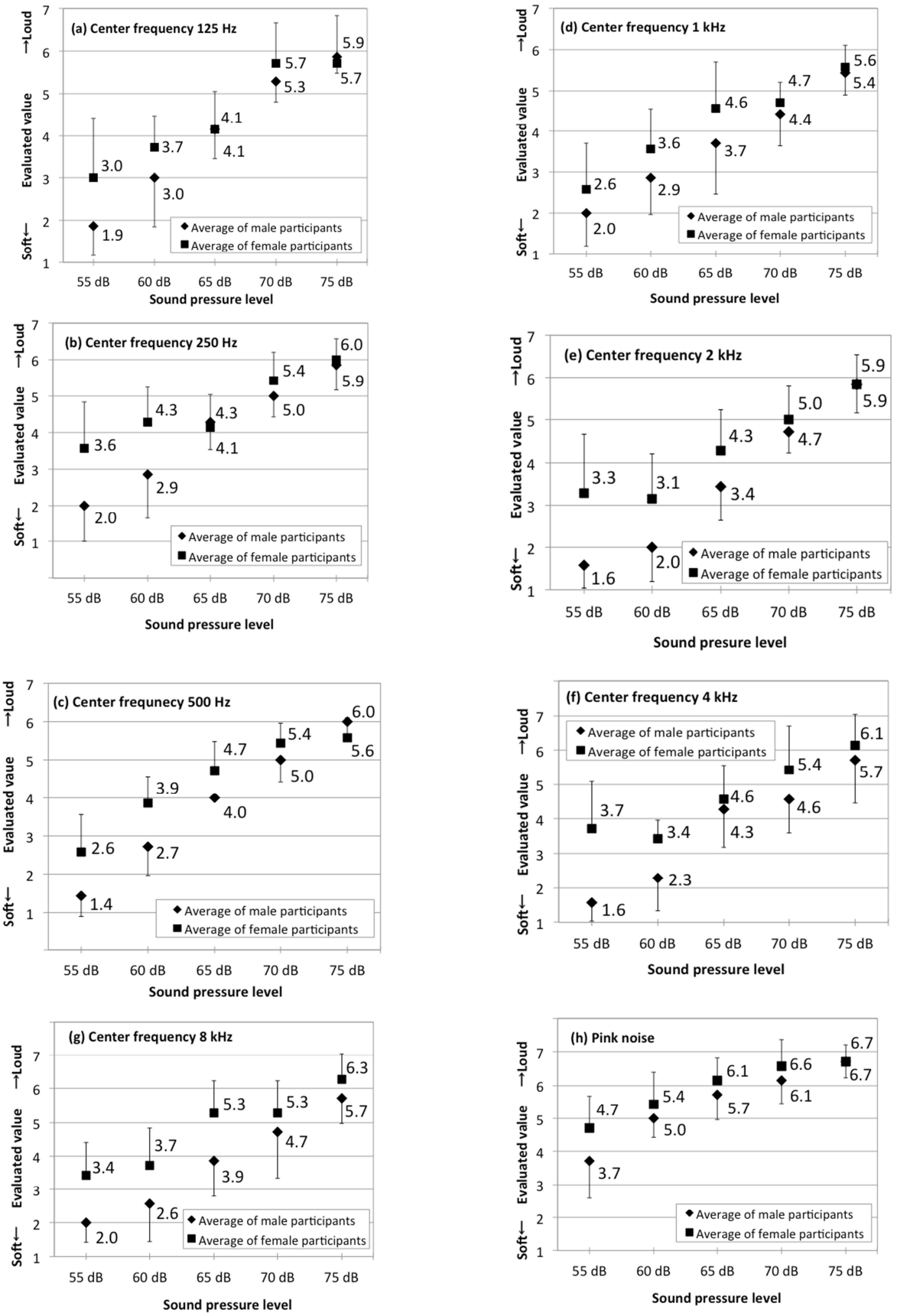


Figure 1: The rating scores for reproduced band noises of male and female participants.

3. RATING EXPERIMENT OF PERCIEVED LOUDNESS OF BROADBAND NOISE AND NARROWBAND NOISE

According to the measurement experiment of the optimum listening level of music, the optimum listening level for the male participants was higher than that for the female participants. To determine the factor affecting the difference of optimum listening level between men and women, we discussed another different factor between men and women. The previous studies reported that efferent inhibition was relatively less in female than in male (McFadden, 1998) and higher sensitivity of hearing in female than in male (Velle, 1987). From these reports, we could hypothesize that there was the difference of the loudness perception between men and women; women might perceive loudness of sound higher than men. If perceived loudness for sound of the same sound pressure level was different between men and women, the optimum listening level of men would be felt 'louder' by women. Therefore, the female participants might adjust the optimum listening level lower than that of the male participants and the difference of optimum listening level between the male and female participants might be observed. The difference of loudness perception between men and women was supposed to be a factor affecting the difference of the optimum listening level of music between men and women. To confirm this assumption, the rating experiment of perceived loudness of broadband and various center frequencies' narrowband noise was conducted.

3.1. Experimental condition and procedure

Pink noise and 1/3 octave band noises of 8 center frequencies (125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz) were used as the stimuli. The duration of each stimulus was 15 s and the sound pressure levels were adjusted to 55 dB, 60 dB, 65 dB, 70 dB and 75 dB. These sound pressure levels were determined from the distribution range of the optimum listening level of music obtained in the present study and the previous study (Hamamura and Iwamiya, 2013). The stimuli were reproduced from a loudspeaker positioned at 2 m in front of the participants. The participants rated the loudness of the stimuli using the rating scale from 1 to 7 (1: very soft, 2: considerably soft, 3: slightly soft, 4: neither, 5: slightly loud, 6: considerably loud, 7: very loud) after listening to each stimulus.

The experiment was conducted in a semi-anechoic chamber. The participants were 14 students of Kyushu University (7 males and 7 females), aged 21 to 32. All of them had normal hearing and there was no significant difference in hearing ability between the male and female participants.

3.2. Results

The average evaluated values on loudness and their standard deviations for each stimulus were shown in Fig. 1. The female participants tended to rate higher loudness scores than those of the male participants for all the stimuli of the same sound pressure level. The three-way ANOVA was applied to the obtained data. As a result, the effect of sound pressure level ($F(4,508) = 257.91$, $p < 0.01$), that of the kind of stimuli ($F(7,508) = 29.666$, $p < 0.01$) and that of gender ($F(1,508) = 87.362$, $p < 0.01$) were statistically significant. In addition, the interaction between sound pressure level and gender was also statistically significant ($F(4,508) = 8.707$, $p < 0.01$).

The effect of sound pressure level means that when the sound pressure levels were louder, the evaluated values on loudness were also increased. Concerning the effect of the kind of stimuli, multiple comparison procedure was applied to the data. The evaluated values for pink noise were higher than those of other narrowband noises. This tendency was due to the fact that we perceived louder for broadband sound than narrowband sound when the total acoustic energy was the same (Zwicker and Fastl, 1990). The effect of gender means the obtained evaluated values for the stimuli of the same sound pressure level were statistically different between the male and female participants. The female participants rated higher loudness scores than the male participants.

The interaction between sound pressure level and gender was based on the difference of evaluation range between the male and female participants. The male participants rated 1 or 2 for 55 dB stimuli while the female participants rated more than 3 for the same stimuli. The loudness evaluation scores of the male participants distributed from 1 to 7 for the sound stimuli from 55 dB to 75 dB while those of the female participants were from 3 to 7.

4. CONCLUSIONS

The difference of optimum listening level of music between men and women was observed. The male participants adjusted higher optimum listening level than that of the female participants. In addition, intra-individual variations of the optimum listening levels for each participant for each stimulus was smaller than the difference of optimum listening level between the male and female participants. The existence of the difference of optimum listening level between men and women was confirmed. This difference should not be observed coincidentally.

The female participants rated higher loudness scores than those of the male participants for the broadband noise and narrowband noise regardless of center frequency of the same sound pressure level. The difference of loudness perception between men and women existed irrespective of the frequency region. Such difference of loudness perception of sound between men and women is a factor affecting the difference of optimum listening level between men and women.

Because of the difference on the loudness perception based on the gender, the female participants might feel 'louder' for the optimum listening level for the male participants. The female participants might feel 'optimum' when the sound pressure level was lower than the optimum listening level for the male participants. Therefore, the optimum listening level of the female participants was lower than that of the male participants

5. REFERENCES

Barrett, D. E. and Hodgets, D. A. (1995). Music Loudness Preference of Middle School and College Students," Texas Music Education Research, 1-6.

Fligor, B. J. and Ives, T. E. (2006). Dose Earphone Type Affect Risk for Recreational Noise-induced Hearing Loss?, Paper presented at the NIHL in Children Meeting.

Hamamura, M. and Iwamiya, S. (2013). DIFFERENCE OF THE OPTIMUM LISTENING LEVEL AND THE PERCIEVED LOUDNESS OF REPRODUCED SOUNDS BETWEEN MEN AND WOMEN, Proc. Of Kyushu-Youngnam Join Conference on Acoustics 2013, 18-21.

Hodgetts, W. E., Rieger, J. M., Szarko, R. A. (2007). The Effect of Listening Environment and Earphone Style on Preferred Listening Levels of Normal Hearing Adults Using an MP3 Player, Ear and Hearing, 20 (3), 290 – 297.

McFadden, D. (1998). Sex Differences in the Auditory System, Developmental Neuropsychology, 14, 261-298.

Velle, W. (1987) SEX DIFFERENCE IN SENSORY FUNCTIONS, Perspective in Biology and Medicine, 30 (4), Summer, 490 - 522.

Williams, W. Noise exposure levels from personal stereo use, International Journal of Audiology, 44, 231-236. Zwicker, E. and Fastl, H. Psychoacoustics: Facts and Models, (Spring-Verlag, Berlin, 1990), 185-191.
Zwicker, E. and Fastl, H. Psychoacoustics: Facts and Models, (Spring-Verlag, Berlin, 1990), 185-191.

Keynote Address 4 & Young Researcher Award 2

SEOUL 16:30 – 18:00 Thursday 7 Aug 2014

A THEORY OF THE MUSICALLY SUBLIME

David Huron
School of Music, Ohio State University, USA
Time: 16:30

ABSTRACT

On occasion, the experience of listening to music evokes strong emotions such as those chronicled by Gabrielsson (2011). Among the various strong emotions induced by music, one class might be deemed "sublime" in the sense described by the 18th century philosopher, Edmund Burke. Burke suggested that sublime emotions arise when pleasure is tinged with fear. For example, Burke suggested that the sense of awe evoked by looking into a deep canyon depended on a muted sense of terror mixed with a feeling of delight. Ethological research on fear has identified four classic fear-related behaviors: fight, flight, freeze, and (in social animals) appeasement. Expanding on an argument made in *Sweet Anticipation* (2006), I suggest that music-induced "chills" (frisson), ASMR, awe, and lump-in-throat/weeping, arise from cognitive inhibition of subcortical fear, and are derived from the fight/flight, freeze, and appeasement responses. I suggest that musical passages that evoke such responses depend on acoustical features associated with fear and that contrastive affect renders these experiences highly pleasurable.

1. INTRODUCTION

In his voluminous documentation of strong experiences induced by music, Alf Gabrielsson's (2011) respondents describe a number of responses. These include the appearance of goose bumps with attendant feelings of chills or shivers. In other situations, listeners may feel a constriction of the pharynx commonly described as feeling "choked up." In some cases, tears may well-up in a listener's eyes, or actually flow as overt weeping. In rare cases, listeners note that they are unable to move; that is, their posture becomes frozen. At the same time, respiration may momentarily cease—an experience captured in the phrase "takes your breath away." When experiencing strong feelings, music listeners may sometimes describe a "tingling" feeling.

If we described these physiological responses to ethologists, they would recognize that all of these behaviors are associated with the experience of *fear*. In his encyclopedic volume on the physiology and psychology of fear, Isaac Marks (1987) describes four classic fear responses. If the threat appears manageable, one option is to fight back. If there is little likelihood of prevailing in a fight, a more prudent response may be to run away as quickly as possible (flight). In situations of extreme fear, an animal may simply freeze—with the hope that the danger will pass without incident.

Finally, a fourth option is typically available for social animals (e.g., wolves, chimpanzees, humans). Among social animals, the most common threats come from conspecifics. For these animals, one response to threat is simply to capitulate—to indicate one's surrender. A surrender display usually brings the aggression to an end, but typically at the cost of a loss of social standing for the surrendering individual. In short, the four responses to fear include *fight*, *flight*, *freeze*, and *appeasement*.

Throughout the animal kingdom, *fight*, *flight* and *freeze* responses share many similarities. Among mammals, it is common for fear to cause the hair to become erect (the pilomotor response), especially on the back, head and limbs. This causes the animal to appear bigger, with the potential to make an aggressor animal reconsider whether to attack. In the case of the freeze response, the animal becomes motionless (tonic immobility). This includes breath-holding.

Among humans, Kottler and Montgomery (2001) have suggested that crying is the *appeasement* display. Especially in disputes or arguments (forms of aggression), the altercation typically ends if one of the parties breaks into tears (Lane, 2006). Research has shown that the most common response to tears is for the observer to stop any aggression and to offer altruistic assistance. The behavioral studies are reinforced by physiological research. Shani Gelstein and her colleagues (2011) collected real and fake tears from volunteers induced to cry by watching sad movie scenes. When men were exposed to the smell of actual psychic tears, testosterone levels and libido measures dropped significantly—consistent with a change of behavior from possible aggression to compassion. As with other animals, the cost incurred by crying is the loss of social stature.

Independent of the physiological resemblance of strong emotional responses in music to fear-related responses in the animal kingdom, a philosophical tradition in aesthetics also suggests a relationship to fear. In his *Philosophical Enquiry into the Sublime and Beautiful* (1757), the Irish political theorist and philosopher, Edmund Burke, argued that "sublime" emotions always include some element of fear. The experience of "awe," might arise when standing on the edge of a cliff, when observing a caged wild animal, or when observing a building on fire. For Burke, "awe" is wonder tinged with fear.

We will expand on Burke's thesis shortly, however there are two further foundational concepts that need to be discussed. The first concept might be dubbed "two brains;" the second concept I have dubbed *contrastive affect* (Huron, 2006).

2. SYSTEM 1 AND SYSTEM 2

Studies of brain functioning suggests a useful broad division between slow cortical and fast subcortical responses. This division is well documented in Joseph LeDoux's dual-pathway analysis of fear processing (LeDoux, 1996), however many other neuroanatomical studies are consistent with this view. Pertinent experiments are reviewed in Daniel Kahneman (2011) who has generalized this distinction and refers to them as System 1 and System 2. System 1 is a fast, instinctive, automatic, defensive, and largely unconscious process that is associated with subcortical structures in the brain. System 2 is a slow, conscious, reflective system that is associated with cortical brain regions. System 1 is always quick to

respond to stimuli, relying on conservative assumptions that are not always appropriate. Often, the assessments arising from System 2 generates an inhibitory signal—belatedly damping System 1 responses. At the same time, System 1 strongly influences and guides System 2, so that conscious, strategic behaviors ensue that better serve the motivations of System 1.

As documented in LeDoux's (1996) work on fear, the auditory system exhibits a dual pathway in which sounds are processed using both quick-and-dirty heuristics, and more rational-cognitive appraisals. The two systems are continuously at work when listening to sounds (see also Huron, 2006; chapter 2).

3. CONTRASTIVE AFFECT

Feeling states appear to be strongly influenced by contrast. If we initially feel bad and then feel good, the good feeling tends to be stronger than if the good experience occurred without the preceding bad feeling. Conversely, if we initially feel good and then feel bad, the bad feeling tends to feel worse (e.g. McGraw, 1999). This 'contrastive valence' may account for the difference between pleasant and unpleasant goose bumps. When we are unexpectedly touched by a stranger, our reaction is negative. But when the unexpected touch comes from a prospective lover, the initial shock of surprise is quickly displaced by a highly positive appraisal with a memorably large contrast in feelings from bad to good. Wrestling with a difficult problem, an unexpected moment of insight replaces a period of stressful rumination with a sudden feeling of achievement. The terror of riding a rollercoaster is held at bay by the cortical conviction that 'I'm not going to die'. Certain sounds may evoke an automatic fearful response. But a cortical appraisal might concludes that 'it's just music.'

4. FEARFUL SOUNDS

This model of sublime emotions was first proposed in Huron (2006), and elaborated in Huron (2008) and Huron and Margulis (2010). A similar theory based on visual art has been proposed by Eskine *et al*. (2012). Specifically, it was proposed that certain sounds activate subcortical "fear" circuits, that cortical processes assess the situation as positive or inconsequential—and so react by inhibiting the subcortical feelings of fear. The pleasure associated with these experiences arises due to contrastive affect between a negative state (System 1) displaced by neutral or positive states (System 2).

This theory raises the question of what sounds might be expected to evoke fear in a listener? In the context of the frisson experience, Sloboda (1991) was the first to describe the music-related correlates. Huron (2006) analyzed a larger sample of musical passages that induced frisson in listeners: seven common features were identified: (i) loud sounds, (ii) approaching sounds, (iii) low sounds, (iv) high volume (many sound sources), (v) scream-like sounds, (vi) close sounds, and (vii) unexpected sounds.

Loud sounds might be expected to evoke fear because they represent large energy events that might lead to injury. Sounds indicating approach (especially rapid approach) are worthy of attention and may indicate attack. The most characteristic acoustic change associated with approach is increasing sound intensity—which musicians refer to as crescendo. With regard to low sounds, one of the best generalizations one can make about acoustics is that large masses or volumes vibrate at a lower frequency than small masses or volumes. Throughout the animal kingdom, threat is associated with appearing large (Bolinger, 1964). Music transposed lower in pitch is perceived as less polite and more aggressive (Huron, Precoda & Kinney, 2006).

The term "volume" was coined by psychologist S.S. Stevens in the 1930s. The word was hijacked by audio equipment manufacturers who considered it better for marketing purposes than *intensity*, *amplitude*, or *loudness*. Unfortunately, this misuse of Stevens' term has obscured an important concept in auditory perception. Stevens himself conceived of volume as a basic subjective experience of auditory "size" distinct from loudness. For Stevens, increasing the number and dispersion of sound sources (as in surround sound, or when augmenting a single violin to a violin section) is correlated with an increase in subjective volume. Multiple sound sources can raise arousal levels and be deemed threatening simply due to the challenges of tracking multiple sound sources, or the connotation of multiple actors in the world, each requiring some degree of vigilance.

The human scream displays a disproportionate amount of energy in the broad region between 1,000 and 6,000 Hz (centered around 3kHz) where human hearing is best. This is true of screams produced by both males and females. The scream resonance can be traced to involuntary movement of the ventricular folds (Beeman, 2005; Sundberg, 1987). The result is a characteristic 'scream' resonance—the human-generated sound that can be heard at the greatest distance. Sundberg (1972, 1977) showed that the so-called "singers formant" produced by opera singers emulates the effect of the ventricular folds by drawing the larynx downward. Once again, especially strong resonances are produced in the region around 3kHz. Apart from opera singers, screaming is commonly found in many popular music styles (including heavy metal, punk, thrash, and screamo). Isolated passages of scream-singing can also be found in many recordings of blues and rock. Moreover, the well-known sound of the distorted electric guitar tends to produce large amounts of energy near 3kHz.

Another source of fear is close or proximate sounds. The sound of a buzzing fly is little cause for concern when the fly is across a room. However, if the fly is near our face, it is difficult not to experience some degree of alarm. Of course there are excellent reasons why an animal should be more concerned about very close sound sources compared with distant sound sources.

A final source of auditory induced fear can be traced to unexpected sounds. Large portions of the brain are dedicated to predicting or anticipating future events. The survival value of accurate expectation is readily apparent. Conversely, the failure to anticipate future events is inherently dangerous. Failures of auditory expectation span a wide range of circumstances (see Huron, 2006 for a detailed exposition). This can include unexpected changes of harmony, abrupt modulations, unanticipated changes of instrumentation, changes of tempo, meter or rhythm, abrupt dynamic shifts, unexpected accents, and so on. Unexpected sounds include chaotic or confusing pitch patterns, unanticipated source locations or trajectories, as well as abrupt changes of texture (including moving from many sounds to one sound, or from a loud texture to a quiet texture).

5. ASMR

In recent years, a new phenomenon has appeared on the worldwide web: popular amateur videos aim to create what fans commonly describe as a pleasant tingling sensation on the surface of the head or scalp, inducing a trance-like state of blissful relaxation. The videos are created using a stereo microphone and involve such activities as a person cutting someone's hair, putting on makeup, drinking a soda, folding towels, sorting through papers, and innumerable other apparently mundane activities. The phenomenon has been given the impressive-sounding name *autonomous sensory meridian response* (ASMR) by aficionados.

The videos share a number of features in common. Most important is a feeling of intimacy. The videos are shot in close-up, with a young person addressing the camera in a highly personal way. The more successful ASMR videos feature attractive young men or women. The voice is typically whispered and eye-contact is commonly made directly with the camera. The videos are intended to be watched with headphones, so in the case of a faux haircut, viewers hear the clipping sounds as though experiencing the haircut themselves. ASMR videos often involve whispering very close to a microphone, mouth sounds such as lip smacking, rustling, soft scratching, rubbing, tapping, or brushing sounds. Close intimate attention is paid to the viewer. At the time of writing, a Bing web search for ASMR returned some 4.5 million hits. The most popular ASMR YouTube videos have over a million views.

ASMR fans claim that the response differs from frisson, however the symptoms described seem to be indistinguishable from frisson. What differs is principally the evoking stimulus. ASMR seems to focus on extreme intimacy and proximity. Recall that very close sounds are linked to alarm and fear. This suggests that ASMR fits within the sublime emotion theory offered here, where a subcortically alarming sound is offset by the cognitive appraisal that the sounds are benign and inoffensive. In contrast to the typically frisson experience, the sounds are very quiet, and so lead to much greater relaxation than evident in common frisson experiences.

6. FEARFUL STIMULI AND RESPONSES

By way of summary, many of the responses associated with strong musically-evoked experiences can be linked to various fear responses. The goose bumps (pilomotor response) is a classic fear-related reaction. The appearance of goose bumps is found in both frisson and ASMR experiences. Breath-holding and frozen posture (tonic immobility) are classic fear-related responses linked to the freeze reaction. These behaviors are symptomatic of the experiences people described as “awe.” Constricted pharynx (“choked up”) and tears are classic crying behaviors--a behavior that is regarded as the human appeasement display. Once again, this is a response that can be traced to stress and fear.

Independent of the resemblance of these behaviors to fear responses, when the evoking music is analyzed, the distinguishing acoustical features can all be plausibly linked to fear. In short, both the stimuli and responses are consistent with the role of fear. Nevertheless, despite the resemblance to fear, in the context of music listening, each of these behaviors is phenomenologically pleasurable.

7. CONCLUSION

On the 22nd of December 1849, the 28-year-old Russian writer, Fyodor Dostoyevsky stood in front of a firing squad on the orders of Czar Nicolas I. It was a mock execution. Many are traumatized by a mock execution, but for Dostoyevsky, the experience was one of sheer ecstasy. The sense of relief was, for Dostoyevsky, a full-fledged spiritual experience.

The theory proposed here expands on the insight first offered by Burke (1757) that sublime experiences appear to involve an element of fear. Specifically, it is argued that the pleasure associated with sublime experiences arises from subcortical (System 1) fears that are inhibited by cortical assessments (System 2), resulting in negative-to-positive contrastive affect. Not all fear responses are the same. Depending on the situation and the magnitude of the fear, physiological responses may most resemble fight, flight, freeze, or appeasement.

Physiological responses such as the goose bumps characteristic of frisson are consistent with fear-engendered efforts to appear bigger. Listener reports of frozen posture and breath-holding are consistent with tonic immobility consistent with the freeze response. The constriction of the pharynx (“choked up”) and incipient or actual tears are consistent with the human appeasement display—crying.

Although laughter at first appears to be an exception, an analysis of laughter episodes suggests otherwise. Elsewhere I have proposed a theory of laughter that is a variant of Vilayanur Ramachadran’s *false alarm theory*. Laughter too, it appears, tends to be evoked when a fear-inducing situation is deemed inconsequential--consistent with pleasure arising from a negative-to-positive contrastive affect.

Both laughter and crying share a vocalized punctuated exhaling that resembles the pant-laughter of non-human primates (Provine, 2000). Elsewhere I have proposed that, both laughter and crying evolved from a single earlier form (“pexhalation”), analogous to pant-laughter in other primates. In most social mammals, a single signal is used both as a sign of appeasement and to solicit play behaviors. For many social mammals, this display involves either voiced or unvoiced panting. However, in humans, I have argued that this single signal (pexhalation) split into two more specialized signals: one as an appeasement display (crying), and a second as a play solicitation display (laughter). The common history for both lies in responses to fear.

In an analysis of musical passages evoking one or another sublime experience, acoustic features associated with fear predominant, including loud sounds, approaching sounds, low sounds, high volume (numerous sound sources), scream-like sounds, close sounds, and unexpected sounds.

In this paper I have proposed that there is an underlying commonality between frisson, awe, pleasurable crying (lacrimission), laughter, and ASMR. In each case, I have proposed that the pleasure arises from contrastive affect where cortical neutral or positive assessment inhibits subcortical fear or alarm.

Neurological tests of this theory are scant. Consistent with the theory, Blood and Zatorre (2001) observed a marked decrease in amygdala activity in response to musically induced frisson. However, due to the poor temporal resolution of PET, the con- jecture that decreased amygdala activity is preceded by a short burst of activity remains untested. However, the status of this admittedly ambitious theory will be determined by future empirical tests.

8. REFERENCES

Beeman, W.O. (2005). Making grown men weep. In A. Hobart & B. Kapferer (eds), Aesthetics and performance: The art of rite. (pp. 23-42). New York: Berghahn Books.

Blood, A.J., & Zatorre, R.J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of the National Academy of Sciences, 98, 11818-11823.

Bolinger, D.L. (1964). Intonation across languages. In J.H. Greenberg, C.A. Ferguson & E.A. Moravcsik (eds.), Universals of Human Language, Vol. 2: Phonology. Stanford, CA: Stanford University Press, 471-524.

Burke, E.A. (1757/1759). A Philosophical Inquiry into the Origin of our Ideas of the Sublime and Beautiful. London: Dodsley.

Eskine, Kendall J., Kacinik, Natalie A, Prinz, Jesse J., (2012). Stirring images: Fear, not happiness or arousal, makes art more sublime. Emotion, February 6, 2012, DOI: 10.1037/a0027200.

Gabrielsson, A. (2011). Strong Experiences in Music: Music is much more than just music. Oxford: Oxford University Press.

Gabrielsson, A., & Lindström, S. (1993). On strong experiences of music. Musikpsychologie: Jahrbuch der deutschen Gesellschaft für Musikpsychologie, 10, 118–139.
Gelstein, S., Yeshurun, Y., Rozenkrantz, L. Shushan, S., Frumin, I., Roth, Y., & Sobel, N. (2011). Human tears contain a chemosignal. Science, 331, 226-230.
Huron, D. (2004). Music-engendered laughter: An analysis of humor devices in PDQ Bach. In S. Lipscomb, R. Ashley, R.O. Gjerdingen, & P. Webster (editors), Proceedings of the 8th International Conference on Music Perception and Cognition. Evanston, Illinois, 2004, 700-704.
Huron, D. (2006). Sweet Anticipation: Music and the Psychology of Expectation. Cambridge, Massachusetts: MIT Press.
Huron, D. (2008). Four sublime musical emotions: Frisson, awe, laughter and weeping. Music and the Brain Symposium, Duke Institute for Brain Sciences, Duke University, Durham, North Carolina (November 13, 2008).
Huron, D., & Margulis, E.H. (2010). Music, Expectation and Frisson. In P. Juslin & J. Sloboda (editors) Handbook of Music and Emotion: Theory, Research, Applications, 2nd edition. Oxford: Oxford University Press, 575-604.
Huron, D., Kinney, D. & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. Empirical Musicology Review, 1, 170-177.
Kahneman, D. (2011). Thinking Fast and Slow. New York: Macmillan.
Kottler, J.A. & Montgomery, M.J. (2001). Theories of Crying. In: A.J.J.M. Vingerhoets & R. R. Cornelius (editors). Adult Crying: A Biopsychosocial Approach. Hove, East Sussex, UK: Brunner-Routledge, 1-7.
Lane, C.J. (2006). Evolution of Gender Differences in Adult Crying. PhD dissertation, University of Texas at Arlington.
LeDoux, J. (1996). The emotional brain. New York: Simon & Schuster.
Marks, I.M. (1987). Fears, Phobias, and Rituals: Panic, Anxiety, and Their Disorders. Oxford: Oxford University Press.
McGraw, A.P. (1999). Expectations and emotions in sports. Masters' thesis. Psychology Department, Ohio State University.
Provine, R.R. (2000). Laughter: A Scientific Investigation. New York: Viking.
Sloboda, J.A. (1991). Music structure and emotional response: Some empirical findings. Psychology of Music, 19, 110-120.

Young Researcher Award 2

SIMULATION OF ONE'S OWN VOICE IN A TWO-PARAMETER MODEL

Sook Young Won[1], Jonathan Berger[1], Malcolm Slaney[2]
[1]Stanford University, USA, [2]Microsoft Research, USA
Time: 17:30

ABSTRACT

It is well known that people often are uncomfortable while hearing their recorded singing and speaking voice. This unfamiliarity with the recorded voice, compared to normal hearing, is due to a different transmission mechanism; listening to one's recorded voice only involves a single air-conduction pathway, whereas the voice we hear when we sing and speak is largely due to a bone-conduction pathway. Despite the well-known phenomenon, one's own hearing has received less attention among researchers since it is a very complex process involving multiple paths from vocal cords to hearing sensation. Furthermore, we are studying the perception of living humans, thus adding more difficulty to proceed mechanical studies because of an ethical reason.

In this study, we aim to measure one's own hearing through a perceptual experiment using a graphical equalizer. We assume that if a subject matches a self-hearing and a hearing of recorded voice by altering slider levels on the equalizer, we can determine spectral characteristics of bone-conduction sound.

First, we design an equalizer consisting of a set of peak and shelf filters for eight frequency bands. Then, we conduct two experiments with different groups as asking participants to find the best fit to their own singing and speech voices by processing their recorded voice on the equalizer.

We estimate transfer functions from air conduction to one's own hearing for both singing and speaking voices based on the chosen equalizer settings. We observe that the transfer functions intra subject are relatively consistent and features mostly band-pass filters, broadly amplifying around 300 Hz to 1200 Hz. Moreover, the averaged transfer functions among subjects also present relatively high degree of similarity regardless of gender and experience level of singing. Finally, we successfully derive a two-parameter model of self-hearing as proceeding experimental data simplification and a validation experiment.

1. INTRODUCTION

When you listen to your own voice, the sound produced by your lungs and vocal folds is delivered to hearing organs through multiple pathways, including the air and bones. However, you are the only person who hears the bone-conducted sound since others only can hear the air-conducted part of the voice. The missing bone-conducted sound is unexpected when you hear a recording of your voice. Figure 1 shows the hearing pathway; a dashed line represents air-conduction (AC) pathway and a solid line indicates bone-conduction (BC) pathway.

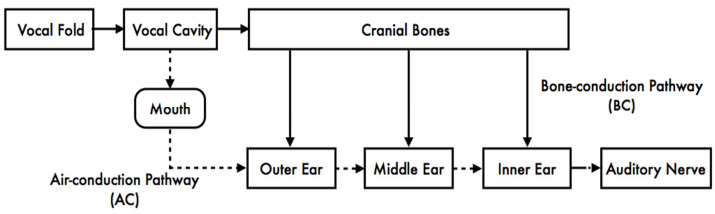


Figure 1: The overview of one's own hearing (von Békésy, 1954; Tonndorf, 1968).

Although the transmission pathways of AC and BC sound are different, both sounds linearly combine at inner ear and excite the basilar membrane similarly (von Békésy, 1932; Stenfelt, 2007). However, since it is impossible to capture the BC sound during vocalization, we cannot just add the BC measurement to AC sound. Therefore, to access the bone-conducted sound, we estimate a transfer function *H* that converts the air-conducted sound to what would be heard as one's own hearing. We do this by extending the equalizer method of Shuster and Durant (2003)

$$\textit{One's own hearing} = H \cdot \textit{AC hearing.} \qquad (1)$$

In this paper, we present details of the equalizer method, analysis of the results, and a simplified model of one's own hearing with the following contributions:

- We describe the implementation of a graphical equalizer software having eight frequency bands optimized for a voice processing. The software use the Cocoa GUI and STK(The Synthesis ToolKit).
- We report results of perceptual experiments with two different groups; amateur singers and professional singers. Then, we explain how to estimate the transfer function *H* based upon the equalizer settings chosen by each subject. The shape of these transfer functions show high degree of consistency intra subject, and are even similar among inter subjects. Overall, the transfer functions feature band-pass filters mostly emphasizing the region from 300 Hz to 1200 Hz.
- Our principle contribution is deriving a model of one's own hearing by decomposing the transfer functions using a Singular Value Decomposition (SVD) and further simplification processes. As a result, the original eight-parameter model, corresponding to the eight frequency bands used in the equalizer experiment, is simplified to a two-parameter model.
- In addition, we describe a validation process and then confirm the model with positive feedback from subjects. Since the model is relatively easy to manipulate and independent of subjects' gender and level of singing experience, we conclude that applying this model on the recorded voice is a feasible way to simulate one's own hearing and also has potential to be a practical application.

2. EQUALIZER EXPERIMENT

We conducted self-perception tests with two different subject groups: amateurs and professional singers. For the purpose of experimental optimization, we designed our own equalizer software in which a recorded voice was altered by peak and shelf filters in real time.

2.1. Experiment Software Design

A graphical parametric equalizer use peaking and shelving filters in order to amplify or attenuate frequencies in the vicinity of a specific center frequency and smoothly connected given gains by interpolation. We adopted second-order peak and shelf filters for the following two reasons. First, these filters possess self-similar shape on a log magnitude scale which agree with psychoacoustic measurement as shown in Figure 2 (a) and (b). Second, their self-similarity enables a linear least-squares optimizer to match a desired dB magnitude in cascaded filters of the equalizer (Abel & Berners, 2004).

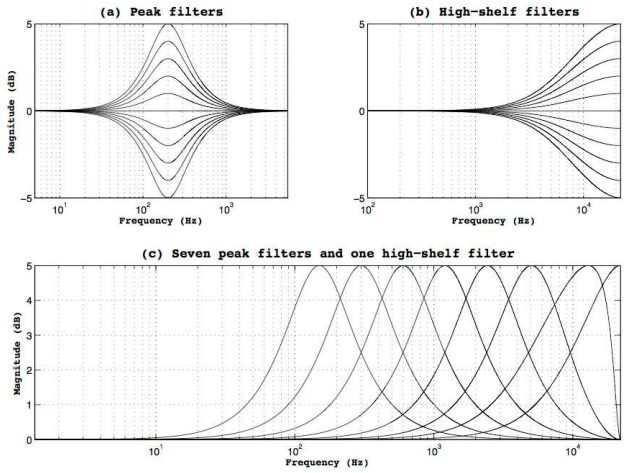


Figure 2: (a) Eleven peak filters and (b) eleven high-shelf filters with a peak gain from -5dB to 5dB with steps of 1dB . (c) Seven peak filters and one high-shelf filter over experimental frequency range with a 5 dB peak.

Since we manipulated speech and singing voice, the experimental equalizer should effectively encompass the vocal range. According to Fastl and Zwicker(2007), the spectrum of speech sounds extends from near 100 Hz to near 7 kHz. Thus, we placed seven second-order peaking filters and one high shelving filter equally over the log-scale frequency axis as shown in Figure 2 (c). The center frequencies of the peak filters were at 150 Hz, 300 Hz, 600 Hz, 1200 Hz, 2400 Hz, 4800 Hz, and 9600 Hz, and the high shelf filter ended at the Nyquist rate – 22050 Hz in our recording setting. Consequently, we had eight frequency bands corresponding to eight sliders in the equalizer. Through several pilot tests, the necessary range of those filters was set from -5dB to 20dB with steps of 1 dB.

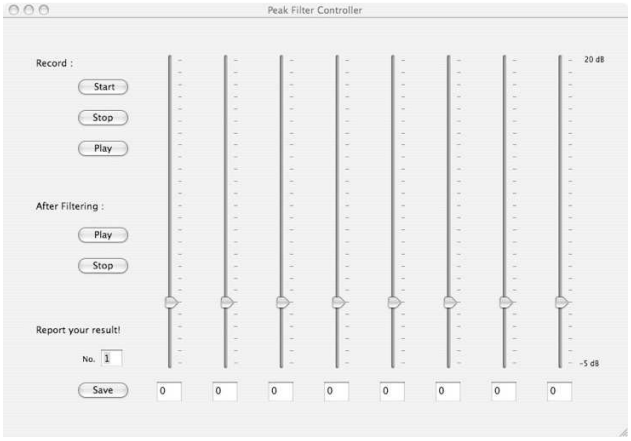


Figure 3: Screenshot of the equalizer experiment software

We designed a graphical user interface to allow a user to easily choose parameters and control the pace of the experiment. The overall design consists of four control parts: (1) recording the subject's voice, (2) interactively selecting a set of different filters, (3) comparing the processed sound resulting from the filter setting selections to one's own voice and (4) finally, saving the selected settings and sound files for the later analyze.

The usage of the experimental software is as follows. A subject first records his/her voice using **Start** and **Stop** buttons in a '**Record**' panel on top-left of the interface. A third button in the record panel labeled **Play** allows for playback of the last recording. Eight sliders in the center of the GUI control the peak and high-shelf filters shown in Figure 2. A numeric field below each slider displays a dB peak level of each filter, which the subject can choose based on her/his preference. Instantaneous changes of the sound are heard either by moving the sliders or by typing a number in the value field. In addition, the latest filtered state can be heard at any time by selecting the **Play** button in the panel labeled '**After Filtering**'. Once the subject encounters the filtered sound bearing the closest timberal similarity to the subject's own voice, the subject submits the filter setting by pressing the **Save** button on bottom-left of the software display, saving the subject's choices in a text file for analysis.

2.2. Experiment Setup

There were two subject pools in our study – a group of 8 amateurs and 13 singers with professional training, and the experiments were carried out at two different locations with each group.

We recruited eight participants who studied at Stanford University for the first experiment. There were 4 males and 4 females in their 20's to early 30's. All participants had non-professional musical training with various instruments such as piano and flute, but not professional vocal training. Thus we defined this group as amateur singers. We carried out the first experiment at the listening room located at CCRMA (Center for Computer Research in Music and Acoustics), Stanford University.

After the first experiment, we expanded the experiment with participants having professional vocal training. We expected that the professional singers might be more sensitive to change of their vocal timbre by the filtering process. Thus we recruited 13 undergraduate students in early 20's who majored in classical vocal music at Seoul National University in Republic of Korea. The subject group consists of six sopranos, one mezzo-soprano, four tenors, one baritone, and one bass. All participants received monetary compensation. The second experiment was carried out in an anechoic room located at Applied Acoustics Laboratory in Seoul National University.

We used exactly the same hardware equipment for both experiments. We recorded vocalizations using a Schoeps CMC microphone placed approximately two inches from the lips of subjects. The recorded signal was digitized using a MOTU Traveler and fed to the experimental software running on a MacBook Pro. As described above, subjects applied filters to the recorded signal and confirmed the changes heard through an AKG K240 headphone connected to the audio output of the MOTU Traveler used for D/A conversion.

2.3. Procedure

Participants were asked to sing eight notes over one octave with a vowel 'Ah' and to speak four short sentences including their name, age, living city and a random sentence. The vowel 'Ah' was chosen to maximize clarity of singing voice timber and its spectral change. The range of singing samples was from C3 to C4 for males and from C4 to C5 for females.

As the first step, participants sung and recorded 3 to 4 seconds with a vowel 'Ah' at C3 (for male) or C4 (for female). Right after recording a sample voice, they used the experimental software to find the closest filter setting to their own voice by altering slider levels. To find the best fit, they sung the same note several times to compare the filtered sound and their own hearing. Usually, it took ten to fifteen minutes to find the best filter set for the first sample, and then it got much shorter. The reference singing note were provided in a given order C3, G3, D3, F3, A3, E3, B3, C4 for male subjects and speech samples were placed between sung notes. The references notes were one octave higher for female subjects.

An entire experiment took from forty minutes to an hour, including time for listening to introduction about the experiment, conducting the equalizer experiment, and giving feedback. Participants marked a total of twelve preferred equalizer settings corresponding to eight singing and four speech samples.

2.4. Result and Analysis

Participants' choices on equalizer

As a result of the two experiments, we obtained 21 matrices arranging the result of each subject's preferences. Each matrix consists of 12 rows (number of voice samples) and 8 columns (number of sliders). Table 1 shows an example of the experimental result matrix from a male

subject in the first group. We listed the results by singing pitch from low to high and then speech. Each number is between -5 and 20 and represents the peak gain of filters in dB according to the participants' choice during the equalizer experiment.

Table 1: Peak gains of sliders. An example test result of a male subject 1 from the first equalizer experiment.

Sample Type	Slider Order							
	1	2	3	4	5	6	7	8
D	13	14	12	10	9	8	6	2
D3	14	13	12	10	8	7	6	2
E3	11	13	13	12	10	8	6	3
F3	10	11	13	13	11	9	7	4
G3	11	12	13	13	11	9	7	4
A3	11	11	12	13	12	10	8	4
B3	10	11	12	13	12	10	8	6
C4	7	9	11	12	12	12	11	8
Speech 1	11	12	12	10	8	6	4	2
Speech 2	15	12	11	11	10	9	8	8
Speech 3	12	13	13	12	11	9	7	5
Speech 4	11	12	13	13	11	9	6	5

Estimating a transfer function of one' own hearing

We estimated the overall transfer function via the following process. We calculated the coefficients of eight filters by applying chosen peak gain and then filtered a 3 second-long impulse though a cascade of peak and shelf filters. Finally, the overall filtered impulse was transformed to frequency response by FFT. This gave us a 2048-point estimate of the desired transfer function.

After applying this estimation process for each sample (each row of Table 1), subject 1 had twelve preferred transfer functions. In general, the loudness of a subject's voice varies over the singing range; individuals commonly tend to sing louder as the pitch increases and this affects the absolute level of the chosen filter settings. Therefore, we normalized the transfer functions by subtracting the maximum magnitudes so that all normalized transfer functions had same peak level, 0 dB. Then, we calculated an average of normalized transfer functions for comparison with other subjects' results. As displayed in Figure 4, the normalized transfer functions of subject 1 in group 1 show a high degree of consistency; most transfer functions have peak between 300 Hz to 1200 Hz (Slider no. 2 to no. 4) except one for singing note C4 which has a peak at 2400 Hz, the fifth peak filter. The standard deviation of these transfer functions is 3 dB. In addition, we could not see noticeable influence by pitch of singing or speech.

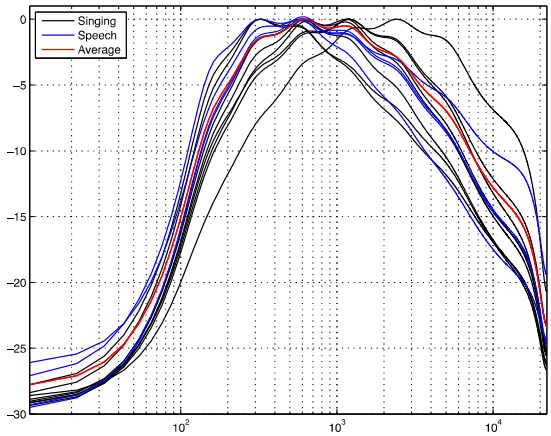


Figure 4: Normalized frequency responses of twelve singing and speech samples and the average of those from a subject 1.

Analysis

We applied this process to all result matrices, and then observed that all participants had self-similar transfer functions of which standard deviations are less than 6 dB intra-subject. Figure 5 displays the eight averaged transfer functions for group 1 and thirteen averaged transfer functions for group 2 separately. In group 1, most transfer functions have peaks at 1200 Hz (the 4th slider) while many of the peaks are at 600 Hz (the 3rd slider) in group 2. Except the two outliers in group 1, all transfer functions in both groups feature a broad band-pass filters with a strong emphasis from 300 Hz to 1200 Hz – some transfer functions had even broader bandwidth extended as high around 2400 Hz. Furthermore, all averaged transfer functions strongly resemble each other. In order to characterize the experimental results, we plot transfer functions by gender per experimental group. However, there is no clear difference caused by gender and level of singing experience.

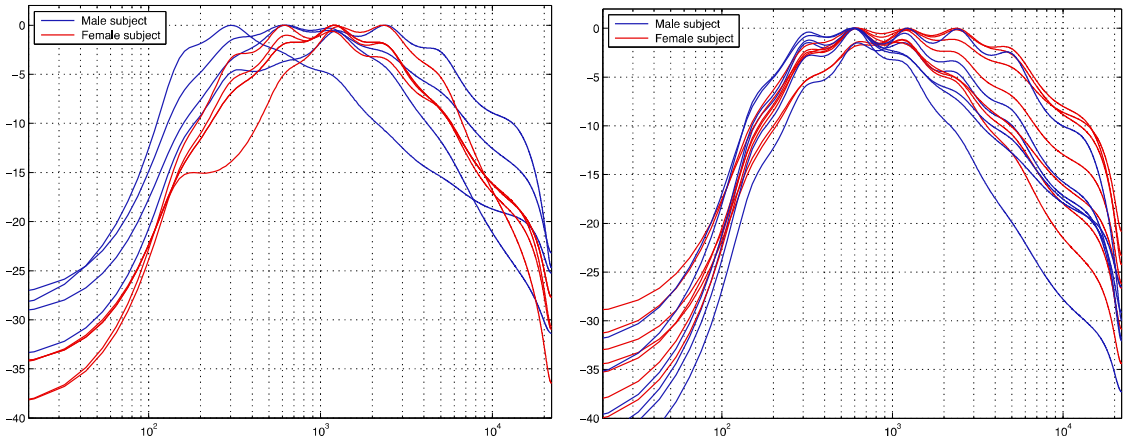


Figure 5: Averaged transfer functions of equalizer experiments.

3. MODELING

3.1. Singular Value Decomposition

We used a singular value decomposition (SVD) to find a low dimensional approximation of the measured transfer function data. The SVD takes a matrix *A* and decomposes the matrix into three matrices; two unitary matrices (*U* and *V*, where *V** is conjugate of *V)* and one diagonal matrix (*S*)

$$A = USV^*. \qquad (2)$$

The total subjects from two experiments were 21. However, we excluded two outliers' results in group 1 for better modeling. Consequently, we considered 19 averaged transfer functions, which transformed air conduction recordings to one's own hearing, into a matrix *A* (2048×19). After applying the SVD in Matlab, we obtained three matrices representing *filter shapes, filter weights* and *individual differences* corresponding to *U* (2048×19), *S* (19×19), and *V* (19×19) respectively.

Advanced Modeling

The first two diagonal values of the filter weight matrix *S* (blue and green circles in Figure 6. (b)) are significantly higher than the other 17 filter weights. By approximating the full transfer functions with the first two singular vectors, we produced a matrix *A′* having high accuracy and reduced dimensionality. Since the differences between matrix *A* and *A′* are relatively small – mostly within -2 dB and 2 dB, we wished to develop a model to simulate one's own hearing with these simplified matrices. We performed several pilot tests with the two SVD filters (the first and second eigenvectors in green and blue lines in Figure 6 (a)). However, some participants complained that it was confusing to use the two filters affecting on the whole frequency range.

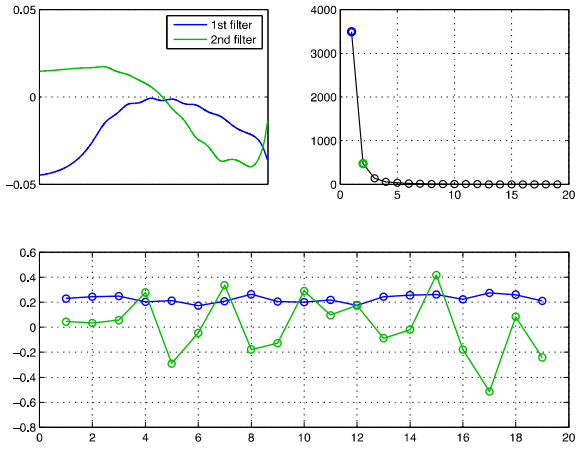


Figure 6: Decomposed and simplified three matrices by SVD. (a) U′ - frequency responses of the first two filters. (b) S - the original diagonal matrix representing filter weight. (c) V′ - individual weight differences of the first two filters.

In order to avoid this side effect of the model with the two SVD filters, we designed a one-parameter model. We set a fixed weight for the first SVD filter since its individual difference (Figure 6 (c)) was relatively small compared to its weight. Then, the model had only one parameter that controlled the weight of the second SVD filter. Another pilot test proved that this one parameter model was simple and easy to control. However, it lost the ability to alter vocal timbre as much as subjects wanted. Therefore, we considered a third model that aimed to have balance between easy controls and flexible simulation.

Our subjects in pilot tests told us that altering the transfer functions by frequency was easiest. Thus, we separated the second filter into two filters at 864 Hz where the filter crosses zero dB. Consequently, the model *M* has one constant filter and two variable filters as shown in equation 3.

$$M = Filter\ 1 + \alpha \cdot Filter\ 2 + \beta \cdot Filter\ 3. \quad (3)$$

During another trials with the model *M*, we finally had positive feedbacks from subjects. Therefore, we decided to perform a validation experiment with this two-parameter model.

4. VALIDATION EXPERIMENT

We recruited another 14 participants, consisting of 7 males and 7 females who studied at Stanford University. Among them, 9 participants had musical singing experience, either musical theatre singing or opera singing. The experiment was carried out in the listening room at CCRMA, as in the first equalizer experiment. We also used the same hardware equipment for recording and listening to the voice.

4.1. Experiment Software Design

We modified the previous experimental software by placing two sliders, corresponding to filter 2 and filter 3, instead of eight sliders manipulating eight frequency equalizer bands. The left panel of the validation experiment software – the six buttons for recording voice, listening unfiltered and filtered voices, and saving the choices – was exactly same as the original equalizer experiment software. The filter weight α for filter 2 are from -2 to 2 with steps of 0.2, and β for filter 3 are from -0.5 to 0.5 with steps of 0.1. The experimental range and step size for filter 2 and filter 3 were tuned through several pilot tests. Figure 7 displays the three filters used in the software to simulate one's own voice.

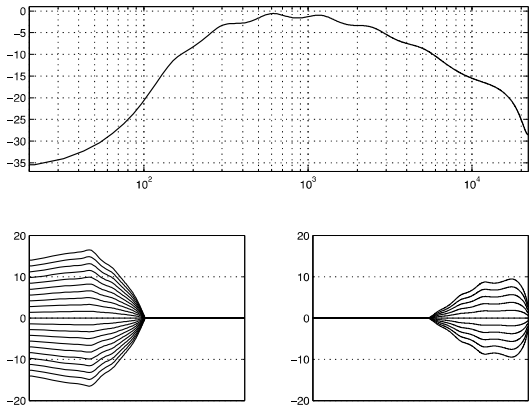


Figure 7: Three filters used in a validation experiment.

4.2. Procedure

Participants recorded three singing and three speech samples of their own choice. Similar to the previous experiments, they attempted to find the best match of their imagined self hearing by adjusting two sliders that controlled the filters applied to their recorded voices. Additionally, they were asked to provide proximity score about their selected filtered voice compared to their own hearing, where the air conduction hearing was zero and their own hearing (AC + BC) was 100.

4.3. Result and Analysis

We obtained filter weights α and β for each recorded sample, thus there were six measurements per participant. The averaged transfer functions from each participant are presented in Figure 8. As expected, the transfer functions are close to each other, although subjects do not know the details of filters processed by the software. Compared to the estimated transfer functions from the original experiments, the peaks of average transfer functions from the validation experiment are lower than the original experimental results perhaps due to the limitation of the controls and predefined filter shapes. More importantly, most participants gave proximity score above 80, up to 95 – one subject answered that his filter selections had 70. Again, we could not find any factors differentiating the transfer functions among subjects such as gender and singing experience.

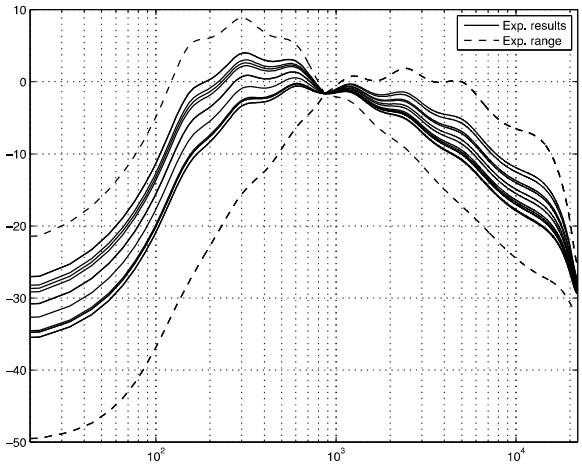


Figure 8: Results of the validation experiment – 14 averaged transfer functions in solid line and the experimental range represented in dashed line.

5. CONCLUSIONS

In our two equalizer experiments, we derived transfer functions *H* from twenty-one subjects and observed that most transfer functions had strong emphasis from 300 Hz to 1200 Hz. This result matches previous studies investigating BC characteristics by measuring the resonant frequencies of a skull. Franke(1956) applied a vibrating piston to a dry skull and found the first resonance to be at 800 Hz. In the same experiment with a skull filled with gelatin, the resonance was reduced to 500 Hz. Later, Pörschmann(2000) estimated the frequency response of bone conduction with a masking experiment and found the amplified region from 700 Hz to 1200 Hz and rapid attenuation above 5 kHz, and these partially accorded with our estimated transfer functions from air conduction to one's own hearing.

Since we obtained a high degree of similarity among the estimated transfer functions inter-subject, we were able to derive a model simulating one's own voice by altering the air-conducted voice with two variable filters. Furthermore, we confirmed the model with a validation experiment. The validation experiment software provided simpler means to simulate one's own voice, however, the accuracy of the transfer functions seems to fell slightly off compared to those processed by the equalizer software.

As a next step, we plan to implement a practical application to allow a user to reproduce his/her own voice in real time. This can be applied in hearing aids for naturally amplifying user's own voice as well as other's voices.

6. REFERENCES

Abel, S. J., & Berners, D. P. (2004). Filter design using second-order peaking and shelving sections. In proceedings of the International Computer Music Conferences, *Miami, FL*.

Fastl, H., & Zwicker, E. (2007). Psychoacoustics: Facts and Models, chapter 9, 259–286. Springer, London, 2nd edition,

Franke, E. K. (1956). Response of the human skull to mechanical vibrations. The Journal of the Acoustical Society of America, 28, *1277–1284*.

Pörschmann, C. (2000). Influences of Bone Conduction and Air Conduction on the Sound of One's Own Voice. Acta Acustica united with Acustica, 86(6), *1038–1045*.

Shuster, L., & Durrant, J. (2003). Toward a better understanding of the perception of self-produced speech. Journal of Communication Disorders, 36(1), *1–11*.

Stenfelt, S. (2007). Simultaneous cancellation of air and bone conduction tones at two frequencies: Extension of the famous experiment by von Békésy. Hearing Research, 225(1-2), *105–116*.

Tonndorf, J. (1968). A new concept of bone conduction. Arch Otolaryngol, 87(6), *595–600*.

von Békésy, G. (1932). Zur theorie des hörens bei der schallaufnahme durch knochenleitung. Annalen der Physik, 13, *111–136*.

von Békésy, G. (1954). Note on the Definition of the Term: Hearing by Bone Conduction. The Journal of the Acoustical Society of America, 26(1), *106–107*.

Poster Session 1

Baekyang Building Lobby
13:00 – 14:00, Monday 4 Aug 2014

[1-01] AUDITORY REORGANIZATION OF GLIDING TONES IN DIFFERENT FREQUENCY RANGES

Gerard Remijn[1], Yushiro Tsubaki[2], Kazuo Ueda[3], Yoshitaka Nakajima[3]
[1]International Education Center, Kyushu University, Japan, [2]Graduate School of Design, Kyushu University, Japan, [3]Department of Human Science / Research Center for Applied Perceptual Science, Kyushu University, Japan

Poster

1. Background: Gliding tones, or tones that change in frequency over time, are natural ingredients of music and speech. In earlier research we have shown that the edges of gliding tones, i.e., their onset and offset, are independent perceptual entities. The edges from physically different gliding carrier sounds can be perceptually integrated into new illusory auditory events.

2. Aim: The aim of the present study was to investigate whether illusory integration of edges from physically different gliding tones depends on the glides' frequency range.

3. Method: We employed the so-called gap transfer stimulus, which consists of a continuous short glide and a longer glide trajectory that is discontinuous in its temporal middle. If the short glide crosses the long glide such that it traverses through the long glide's gap, the gap in the long glide is perceived as if it were in the short glide. This 'gap transfer' is arguably caused by the perceptual reorganization of the gliding tone edges around the gap into new auditory events. In the present study we varied the center frequency of the gap transfer stimuli from 500, to 1000, 3000, and 5000 Hz. Furthermore, we varied the frequency distance between the crossing glides. The glides crossed each other exactly at the center frequency, or were separated by 1, 3, or 5 equivalent rectangular bandwidths (ERBs) relative to the center frequency. Twelve participants judged the perceived (dis-)continuity of the gliding components in the gap transfer stimuli on a rating scale.

4. Results: The listeners' judgments showed that the illusory gap transfer declined as the frequency distance between the long and the short glide increased. The effect of frequency distance on the occurrence of illusory gap transfer, however, strongly varied with stimulus center frequency. As the center frequency of the stimulus increased, the illusory gap transfer became more persistent, and occurred even under conditions where the long and the short glide were separated by more than one ERB.

5. Conclusion: The present results suggest that the temporal structure imposed by the auditory system on gliding sounds can vary with the frequency range of the glides.

[1-02] DISCOVERING PERCEPTUAL CHARACTERISTICS OF MULTICHANNEL MUSIC WITH HEIGHT AMBIENCES FOR JAPANESE LISTENERS

Hidetaka Imamura[1], Sungyoung Kim[2], Atsushi Marui[1], Toru Kamekawa[1], Richard King[3], Wieslaw Woszczyk[3]
[1]Tokyo University of the Arts, Japan, [2]Rochester Institute of Technology, USA, [3]McGill University, Canada

Poster

ABSTRACT

This study identified salient perceptual characteristics of Japanese listeners upon which they conceptualize multichannel-reproduced music with height ambiences and investigated influence of experimental variables including height-loudspeaker position, height signals, musical selection, and ambience level balance. Room ambiences were generated by convolving pre-captured room impulse responses at multiple locations with anechoic music recordings. Trained listeners were asked to evaluate various multichannel music stimuli with height ambiences and rank them. The results indicated that height-loudspeaker position and level balance differentiated reported listeners' rank data. Furthermore, the elicited descriptors showed that listeners ranked high a specific multichannel music that was more enveloping and reverberant.

1. INTRODUCTION

Music and space is closely related and musicians play with the space to elaborate their performances. Acoustic responses of a space have been therefore an essential component in music reproduction even from the very early stage of audio technology. Recording engineers have developed various methods to effectively capture the acoustic responses of a space. However, delivering captured responses has been limited by transmission technologies. As data capacity of real-time transmission increases due to recent digital telecommunication technologies, the use of high resolution and multichannel audio for music reproduction became popular even for consumers. High-resolution audio data can preserve the original tonal quality of recorded music and multichannel audio can present enhanced spatial characteristics.

For instance, a conventional ITU recommended 5-channel audio can deliver auditory images featuring surrounding and enveloping a listener with enhanced localization. The loudspeakers located at the rear-side and the front-center contributed to precise localization and presence [1]. While the current format significantly enhanced spatial realism in the horizontal plane, listeners anticipate similar spatial enhancement in vertical axis. For such vertical enhancement, many institution proposed a new multi-layer audio formats utilizing 9.1-channel [2], 10.2-channel [3], 22.2-channel [4], and more.

The new channels added to the conventional 5-channel introduced greater immersion and presence. For instance, Hamasaki, *et al*. [5] showed that the height channels improved the presence and reality of sound reproduction coupled with Ultra High-Definition (UHD) video contents. Kamekawa showed that height channels contributed to better integrate auditory information with 3D visual contents [6].

Previously the authors have investigated perceptual characteristics of multichannel music, and reported that in multichannel piano music, salient perceptual characteristics were perceived width, perceived sharpness, and perceived bass-tightness [7]. In this

study, we extended previous study to the perceptual characteristics of multichannel-reproduced music when coupled with height room-ambiences radiated from above listeners.

In this research project, we controlled four experimental variables: loudspeaker position of height ambience, contents of height signals, musical selection, and ambience level balance. Furthermore, we collected a set of descriptors that a listener used to express perceptual characteristics associated with variation of height ambiences in a multichannel reproduced music.

2. EXPERIMENT

2.1. Stimuli preparation

The authors utilized two sets of room impulse response (RIR) measurement data to generate multichannel-music sound for the experiment. The RIRs of two venues (having RT60 at 500Hz of 1.40 sec and 2.51 sec respectively) were measured at 32bit / 96kHz resolution using an 8-channel surround microphone array [8]. The array consists of six horizontal microphones and two height microphones. Six microphones were four omni-directional (located at front-left (FL, -45º), front-right (FR, +45º), rear-left (RL, -135º), and rear-right (RR, +135˚)) and two bi-directional microphones (located at front-center (FC, positive lobe facing to 0˚) and rear-center (RC, 90º)). Two height microphones were bi-directional located at azimuths of ±90º and elevations of ±45˚ with positive lobes pointing upwards and an inter-microphone spacing of 1 meter. The RIRs were measured at multiple distances and heights of each venue.

Table 1: The list of control variables used in the experiment.

Stimuli variables	
Level balance	Producer A, Producer B
Music Program	Music A, Music B
Acoustic condition	Studio, Church

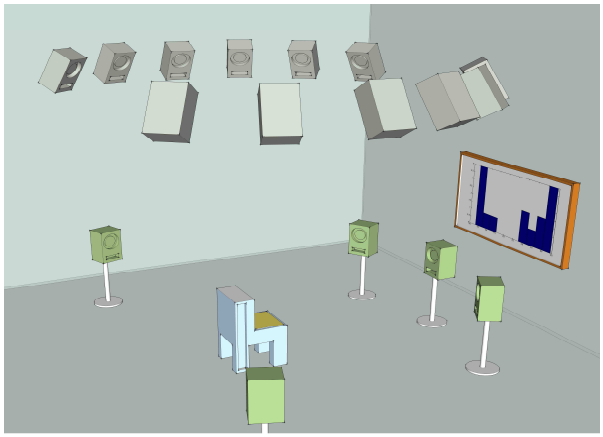


Figure 1: Illustration of the 17 loudspeaker positions (five horizontal and 12 height) for the experiment.

Subsequently we convolved two anechoic music recordings with the RIR data, generating two sets of room ambiences. This gave us four sets of room ambiences (two venues and two music). Among the four sets of room ambiences, two professional recording producers chose nine room ambiences to generate a nine-channel music stimulus, which was reproduced via nine loudspeakers. Horizontal loudspeakers were located at the ITU recommended positions and the height loudspeakers were located at azimuth of ±30º and ±110º, and elevation of +30º, directly above L-, R-, LS-, and RS-loudspeaker of the horizontal plane. This position was a reference point in the subsequent listening experiment. One producer mixed at the A820 (Immersive Presence Lab) of CIRMMT at McGill University and another producer mixed at the Studio B in Senju Campus of Tokyo University of the Arts. Table 1 shows the stimuli combination used in the experiment.

2.2. Reproduction conditions

Listening tests were conducted in the Studio B in Senju Campus of Tokyo University of the Arts where one producer mixed one set of stimuli. The room volume is about 50 m^2, and its reverberation time was 0.34 s (T30 at 500Hz). Five Genelec 8260A loudspeakers were used as horizontal loudspeakers according to the ITU recommendation at listener's ear height (1.2m).

Twelve matched Genelec 8020B loudspeakers were used to generate various reproduction conditions of four height ambiences. These height loudspeakers were located at the elevation of 30˚ and azimuth angle of ±30º, ±50º, ±70º, ±90º, ±110º, and ±130º. Among those 12 loudspeakers, only four loudspeakers were used to reproduce height ambience at a time. While a total of 15 conditions were possible (with 12 symmetrically located loudspeakers), we reduced the conditions to eight by using one front height loudspeaker and one rear and by eliminating the case where the front and rear were adjacent (for instance, ±70º and ±90º were excluded from the experiment). Table 2 shows the elevation and azimuth angles of eight height ambience reproduction conditions. Figure 1 illustrates the positions of 12 loudspeaker used in the experiment.

Table 2: Height-loudspeaker conditions and their positions used in the experiment.

Condition	Elevation angle	Azimuth angles (Front, Rear)
C1	+30˚	±30º, ±90º
C2	+30˚	±30º, ±110º
C3	+30˚	±30º, ±130º
C4	+30˚	±50º, ±90º
C5	+30˚	±50º, ±110º
C6	+30˚	±50º, ±130º
C7	+30˚	±70º, ±110º
C8	+30˚	±70º, ±130º

2.3. Subjects

A total of fourteen subjects participated in the listening experiment. All of them were students and faculty members of the Tokyo University of the Arts, studying and teaching sound recording and psychoacoustics. Although we did not conduct a formal audiometry test, all listeners had no difficulty in both everyday listening and critical listening for music production. The subjects also received the timbral ear training and had experience as a participant for subjective listening tests.

2.4. Evaluation method

In this experiment, listeners participated in total eight listening sessions where they were asked to evaluate perceived impression of multichannel-reproduced music with height ambiences differentiated by height-loudspeaker conditions in two steps.

In the first step, listeners were asked to compare the eight height-loudspeaker conditions and rank them based on the perceived degree of appropriateness using an ordinal scale ranging from 1 (least) to 8 (most). In other words, listeners had to find most appropriate use of the four height-ambience channels in the context of surround music reproduction and provide the corresponding rankings. Listeners could switch freely between the loudspeaker conditions at any time using a custom-made user interface illustrated in Figure 2.

In the second step, listeners made descriptive adjectives associated with their rank data using the Triadic comparison method [9]. They first were asked to choose two most similar height conditions out of three conditions: the highest ranked and lowest ranked in the first step, and the reference condition C2. Subsequently they were asked to provide a descriptive adjective for reason why the chosen pair was similar to each other. Then they described why the left one was different from the similar pair. Since we wanted to understand perceptual characteristics of Japanese listeners, listeners were allowed to describe perceived similarity and/or dissimilarity using Japanese words. Figure 3 show the interface used in the second step.

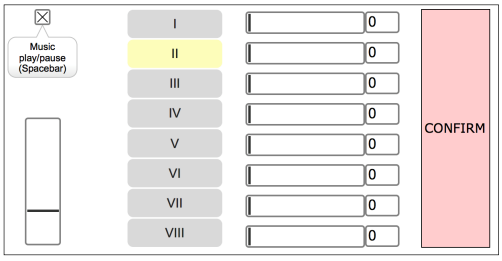


Figure 2: GUI used in the first step

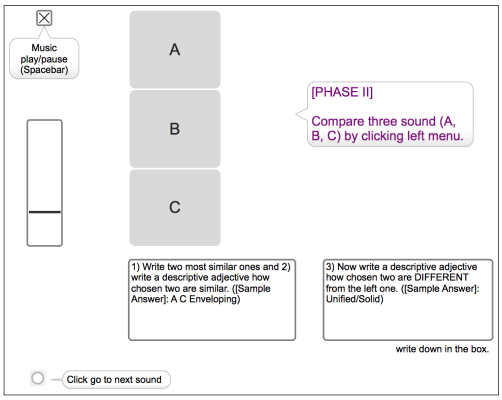


Figure 3: GUI used in the second step

3. RESULTS

Total 112 rankings (two Level balances × two Musical programs × two Acoustic conditions × fourteen listeners) over eight height-loudspeaker conditions were collected through the first step. In the second step, total 244 adjectives were collected, which were related to the perceptual characteristics associated with the most and least appropriate conditions.

3.1. Rankings on the loudspeaker condition

We conducted the Friedman test, equivalent to analysis of variance (ANOVA) for rank data, in order to statistically investigate whether the height-loudspeaker condition had influenced the listeners' rank data or not. The result showed that the rank data were differentiated by the loudspeaker conditions significantly ($p < 0.05$).

Table 3: Rank sums of each loudspeaker conditions.

Speaker Condition	C7	C8	C4	C6	C2	C3	C5	C1
Ranking sum	565	545	525	513	502	477	456	447

Table 3 shows rank sums of each reproduction conditions from C1 to C8 ordered from highest toward lowest. The Wilcoxon signed-rank test with Bonferroni correction was done to see which rank sums of loudspeaker condition is statistically different from the others. The results show that C7 is significantly ranked higher then the group of C5 and C1 ($p < 0.05$). We can state that the listeners in this experiment chose the condition C7 as the most appropriate one, and the C5 and C1 as the least appropriate ones.

3.2. Description analysis

To understand perceptual characteristics corresponding to the stimuli, the authors analyzed the collected descriptors. First, we categorized 244 adjectives into 38 adjectives by grouping similar meaning adjectives together. Then, within each group, we counted the number of occurrences of adjectives on each loudspeaker conditions.

Table 4 shows four most frequent representative descriptors with number of occurrences corresponding to the condition C7, C1 & C5, and the other conditions. We directly translate Japanese descriptors to English descriptors.

We conducted a $\chi 2$ test to statistically observe whether the accumulated occurrence numbers (as categorical data) were significantly different or not. The result showed that frequency of "Reverberance" was significantly different over three groups of reproduction conditions.

Subsequent within-condition analysis showed that listeners mainly described the condition C7 as enveloping and reverberant, which could be regarded as two salient characteristics required for appropriate multichannel music reproduction coupled with height ambiences.

Table 4: Four representative descriptors and their frequency for three groups of height conditions.

Described perceptual characteristics	C7	C2 C3 C4 C6 C8	C1 C5	p-value from χ^2 test
Envelopment	8 / 30	15 / 30	7 / 30	$p = 0.06$
Clarity	4/ 26	16 / 26	6 / 26	$p = 0.90$
Width	2 / 22	18 / 22	2 / 22	$p = 0.15$
Reverberance	6 / 18	6 / 18	6 / 18	$p = 0.01$

4. ANALYSIS OF PHYSICAL PARAMETERS

After analyzing perceptual descriptors, we attempted to discover a physical parameter that could account for the variance of the collected rank data. While the entire physical analysis of room impulse responses and binaurally recorded sound is beyond the scope of this paper, we would like to present a meaningful finding here as an example. We analyzed a binaural signal of the Music A with Church acoustics balanced by the Producer A. The binaural signal was measured using the Bruel & Kjaer Head and Torso Simulator Type 4128-C, which was placed at the listening point. The signal was first divided to 25 1/3-octave bands, and then we calculated each band's Ear Signal Correlation (ESC). Through the stepwise regression, we calculated weighted sum of ESCs that accounted for the variance of rank sum data. As shown in Figure 4, the regressed physical predictor—weighed sum of Ear Signal Correlation (ESC)—covaries with the collected rank sum data (r^2=0.73). While preliminary, this result implies that height channel configuration influenced on similarity between two ear signals, which caused listeners to choose a specific configuration over others. In other words, it might be possible to pre-evaluate a specific loudspeaker configuration in a room and to adjust the configuration that would generate an optimal ear signal difference using the physical predictor. The authors are currently analyzing all stimuli to generate a set of reliable physical predictors that would better account for variance of the control variables.

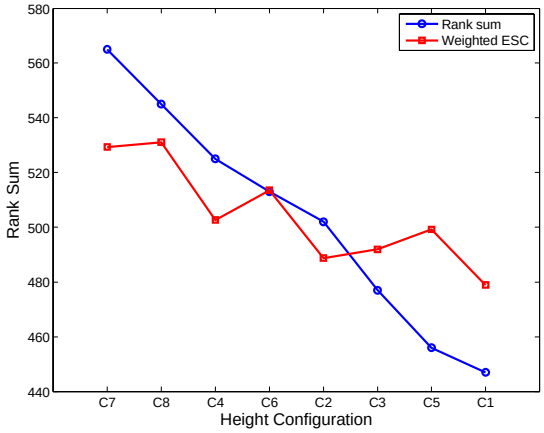


Figure 4: The collected rank sum data and the regressed physical predictor—weighted sum of 25 Ear Signal Correlation (ESC) measured at 1/3-octave bands.

5. DISCUSSION

The main effect (height-loudspeaker position) of the experiment actually consisted of multiple variables. To understand the influence of other variables on the main effect, we conducted additional Friedman tests on the variables listed in Table 1.

The results show that the Music program and Acoustic condition variables did not significantly differentiated collected rank data. The Level balance however influenced on the listeners' rank data. The rank data from the level balance made by "Producer B" were significantly different when analyzed through the Wilcoxon signed-rank test with Bonferroni correction. The subsequent Friedman test (on "Producer B" data) on 56 rankings (4 stimuli times 14 listeners) also showed that the rank data were differentiated significantly ($p < 0.01$). Yet the "Producer A" data did not significantly differentiate the rank data.

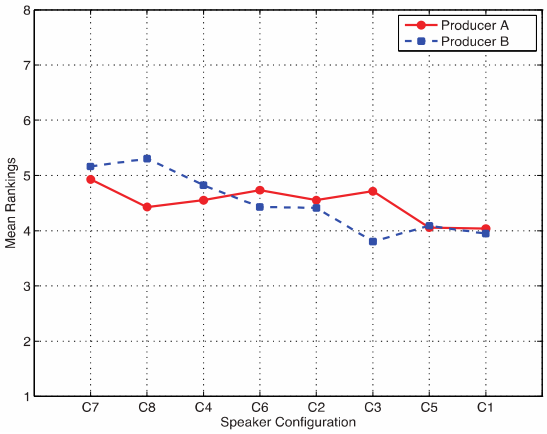


Figure 5: Mean rankings on each loudspeaker conditions of "Producer A" and "Producer B"

This result implicates that an optimally balanced height ambience might not reveal its intended spaciousness when reproduced at other acoustic environments (The Producer A and Producer B mixed in different rooms). In the use of height ambience, the physical room dimension and associated room acoustics could be an important factor, which has to be investigated in-depth. For this, the authors conducted the same experiment at two different locations (Rochester Institute of Technology (RIT) and McGill University) and planed to analyze the influence of the listening room on the discrimination of fine spatial difference associated with the height ambience.

6. CONCLUSIONS

This paper describes an experiment that investigated which height-loudspeaker position would be the most appropriate for multichannel-reproduced music with height ambiences, and that established perceptual characteristics associated with the height loudspeaker conditions.

Rankings over eight height-loudspeaker conditions for total eight multichannel music stimuli were collected through the first step of the experiment. In addition, a total of 244 associated adjectives were collected through the second step of the experiment.

The following results were found for the presented experiment.

1. Height-loudspeaker position of C7 (Elevation angle:+30˚,Azimuth angles:±70º ±110º) was the most appropriate conditions chosen in this experiment. C1 (Elevation angle:+30˚,Azimuth angles:±30º ±90º) and C5 (Elevation angle:+30˚,Azimuth angles: ±50º ±110º) were the least appropriate conditions chosen in this experiment. Other conditions showed no significant difference.

2. Listeners mainly perceived the condition C7 where the height ambiences delivered two salient perceptual characteristics, envelopment and reverberance.
3. A binaural analysis showed that weighted sum of Ear Signal Correlation (ESC) could explain the variance of the collected rank data. While preliminary, this result showed that an appropriate 4-ch height ambience created difference in listeners' ear signal that caused them to choose a specific configuration over others.
4. The Level balance variable appeared to differentiate the listener's rank data. In other words, there was a specific level balance that allowed listeners to better discriminate fine spatial differences, and it might be possible that a specific room acoustics is a reason to create such difference.

7. ACKNOWLEDGEMENT

This study was supported by Yamaha Corporation and the Center for Interdisciplinary Research in Music Media and Technology (CIRMMT).

8. REFERENCES

[1] Francis Rumsey (2001). Spatial Audio, *Focal Press, Oxford, UK.*
[2] Auro3D Technologies (2013). AURO 3D Listening formats, http://www.auro-technologies.com/system/listening-formats.
[3] Kim, et al. *(2010). New 10.2-channel Vertical Surround System (10.2-VSS); Comparison Study of Perceived Audio Quality in Various Multichannel Sound Systems with Height Loudspeakers.* Proc. AES 129th Int. Conv., *San Francisco, USA.*
[4] Hamasaki, et al. *(2005). The 22.2 Multichannel Sound System and Its Application.* Proc. AES 118th Int. Conv., *Barcelona, Spain.*
[5] Hamasaki, et al. *(2006). Effectiveness of Height Information for Reproducing Presence and Reality in Multichannel Audio System.* Proc. AES 120th Int. Conv., *Paris, France.*
[6] Kamekawa, et al. *(2011). Evaluation of spatial impression comparing 2ch stereo, 5ch stereo, and 7ch surround with height channels for 3D imagery.* Proc. AES 130th Int. Conv., *London, UK.*
[7] William L. Martens and Sungyoung Kim. (2007). Verbal Elicitation and Scale Construction for Evaluating Perceptual Differences between Four Multichannel Microphone Techniques. Proc. AES 126th Int. Conv., *Vienna, Austria.*
[8] Woszczyk, et al. *(2009). Selection and preparation of multichannel room impulse responses for interactive low latency rendering of virtual rooms.* Proc. the 16th Int. Cong. on Sound and Vibration, *Poland.*
[9] Leveit, et al. *(1966). Triadic Comparisons of Musical Intervals.* The British Journal of Mathematical and Statistical Psychology, *19, part 2, 163–179.*

[1-03] PERCEPTUAL FUSION AND SIMULTANEITY FOR AUDITORY STIMULI

Satoshi Okazaki[1], Makoto Ichikawa[2]
[1]Graduate School of Humanities and Social Science, Chiba University, Japan, [2]Faculty of Letters, Chiba University, Japan
Poster

ABSTRACT

The perception of simultaneity of two auditory tones has been analyzed in accordance with an implicit assumption that judgment of synchrony is a consequence of a single process. However, this assumption has not been demonstrated by any empirical study. Herewith, we examined the perceptions of relative timing for auditory stimuli in accordance with the assumption that the judgment of synchrony is based on multiple processes. In experiments, we introduced the notion that listeners' judgment of relative timing for two tones involves three categories. First, the two tones are perceptually fused, and the listeners perceive those two tones as a one tone. Second, the two tones are perceptually distinguished, but the listeners perceived them as simultaneous tones. Third, the listeners perceived the two tones as separate and asynchronous tones. In our experiments, participants classified two tones with different SOA (stimulus onset asynchrony) by the use of these three categories. We found that simultaneity perception involves fusional judgment and synchronizing judgment. In addition, we found that the judgment for the perceptual fusion are narrowly distributed around zero SOA while that for the perceptual synchronization are broadly distributed, and that the peak SOA for the perceptual fusion was significantly apart from zero SOA towards positive SOA while that for the perceptual synchronization was around zero SOA. These results suggest that the perception of simultaneity depends not upon a single process, but upon multiple processes.

1. INTRODUCTION

The studies of perceptual simultaneity have been developed with implicit assumption that only a single process underlies the perception of the relative timing for auditory stimuli. However, there have been some disagreements among those previous studies. On the one hand, Parker (1988) examined the perceived simultaneity for two pure tones with various frequency set. She showed that the range of SOA for the two pure tones that are perceived as simultaneous increased with the increment of the frequency distance between the two tones. On the other hand, Wojtczak, Beim, Micheyl, and Oxenham (2012) showed that that range of SOA for simultaneity was constant for varying frequency separation between the two tones with various frequency sets. In addition, Parker (1988) reported that the range of SOA for the perceptual simultaneity was wider than that reported by Wojtczak et al (2012).

There are some studies which propose that multiple processes underlie the perception of relative timing for auditory stimuli. For instance, Hirsh (1974) theoretically claimed that the perception of temporal order is based upon multiple processes, rather than upon a single process. That is, first, listeners perceive a single sound in listening two tones with quite small SOA. Second, listeners perceive a grouped sound in listening two tones with relatively small SOA. Third, listeners perceive successive sounds in listening two tones with large SOA. Noorden (1975) studied auditory stream segregation, and he clarified that there are three categories which account for the perception of stream segregation. Listeners were presented tone sequences (ABAB…) which were composed of two tones with different frequencies. Then, first, listeners perceive two streams (AA… and BB…) for fast frequency alternation of sequences. Second, listeners can perceive either two or one streams depending listeners' attention for relatively slow frequency alternation of sequences. Third, listeners perceive one stream (ABAB…) for two considerably slow frequency alternation of sequences. These multiple processing of the perception of relative timing might cause the discrepancy among the previous studies.

Since simultaneity perception for two tones depends upon the relative timing of those tones, the possibility that multiple processes underlie the perception of simultaneity should be considered in measuring adequately the range of SOA between the tones within which listeners perceive those tones as simultaneous. Moreover, there are three possible categories which concern with the perception of simultaneity. First, listeners perceive a single fused tone for two physically asynchronous tones if SOA between the tones is quite small. Second, listeners perceive separated but simultaneous tones if SOA between those tones is relatively small. Third, listeners perceive asynchronous tones if SOA between those tones is relatively large.

In this study, we examined the possibility that simultaneity perception is based not upon a single process, but upon multiple processes. Conventional simultaneity judgment task had offered listeners only two categories, "synchrony" and "asynchrony". However, as we described above, there should be three possible categories for the perception of simultaneity. In this study, we introduce these three categories as alternatives in simultaneity judgment task for two pure tones. By the use of these categories, listener's simultaneity perception could be measured appropriately.

To test the possibility that simultaneity perception is not a consequence of a single process, we used the frequency intervals which promote perceptual fusion for the two pure tones. There are various frequency intervals which do not cause the perceptual fusion for the two tones. However, if we can demonstrate that the simultaneity perception depends upon not only the synchronizing judgment, but also the fusional judgment by the use of any sets of pure tones with different frequencies, we may claim that the perception of simultaneity is based upon not only a single process, but upon multiple processes.

We measured the range of SOA and peak SOA between two tones with which listeners perceive the two tones as fused into one sound, and as synchronous tones. In addition, we investigated the effect of frequency separation between the tones on the perceptual fusion and perceptual asynchrony for the tones. As we described above, there have been discrepancy among previous studies in considering if frequency separation between the tones affects the range of SOA within which listeners perceive the two tones as simultaneous. Wojtczak et al. (2012) showed that the frequency separation between two tones did not affect the peak SOA with which listeners judged most frequently that the two tones are simultaneous. Although these findings are consequences of the conventional simultaneity judgment task, we analyzed the effect of frequency separation for the fusional processing and synchronizing process separately.

2. METHOD

2.1. Apparatus

The stimuli were presented at 80 dB via Behringer C50A loudspeaker with Scarlett 2i2 audio interface. An Apple Macbook Air laptop computer generated pure tones with 44.1 kHz sampling rate and 16 bit resolution. Participants faced loudspeaker at eye level and with a distance of 41.2 cm. They seated in a KAWAI FKS20-12 sound-proof room. They were instructed for the task by the use of the computer display on the desk.

2.2. Stimuli

We presented two pure tones with different frequencies. That is, the lower frequency was fixed at 262 Hz while the higher frequency was either 393 Hz or 524 Hz. These frequency relations are the perfect fifth and the octave in musical interval. It is known that these simple integer ratio intervals possibly cause perceptual fusion in listing those two tones. Because one of our purposes was to obtain the counterexample to the explanation that simultaneity perception comprises a single process, we used these frequency intervals in order to make fusional process involved in relative timing judgment. The relative onset asynchrony between the two tones was ±0, 2, 4, 7, 14, 27, 52, or 100ms (positive values indicate that the tone with lower frequency preceded the tone with higher frequency). While the two tones began asynchronously, they ended at the same time. The following tone lasted for 300ms. This indicates that the total duration of the two tones was the sum of the asynchrony between the two tones and duration of the following tone (300 ms). Each tone was tapered with a rise-fall time of 15 ms, which was defined by cosine function.

2.3. Procedure

Each condition for the stimulus onset asynchrony between the two tones with different frequencies was presented ten times in random order. That is, each participant performed 300 trials consisted of fifteen SOAs, two frequency sets, and ten repetitions. After each trial, participants were asked to judge whether the two tones were perceptually fused or separated. Only if they judged the tones were separated, they were further asked to judge whether these two tones were simultaneous or not. There was no feedback for the participants' judgment.

2.4. Participants

Five females and six males (including the first author) participated in this experiment. Their ages ranged from 19 to 34 years.

3. RESULTS

We plotted the frequency of each fusional and synchronizing judgment against SOA conditions (Fig.1). Overall simultaneous judgments increased with the decrement of SOA between the two tones. Fusional judgment itself also increased with the decrement of the SOA.

Standard deviation of frequency distribution was used as an index of the range of perceptual fusion and synchronization. A repeated-measures ANOVA with judgment type (fusion or synchronization) and frequency separation as factors found a significant main effect of the judgment type [$F(1,10) = 31.06$, $p < .001$]. However, the main effect of frequency separation [$F(1,10) = 0.02$, $p > .05$], nor the interaction of these two factors [$F(1,10) = 1.34$, $p > .05$] were not significant (Fig.2, left). The mean of frequency distribution (the mean of the product of frequency and SOA) was used as an index of the peak SOA of fusion and synchronization. The peak SOAs of fusion were significantly apart from zero SOA towards positive SOA [262 Hz with 393 Hz: $t(10) = 4.00$, $p < .05$, 262 Hz with 524 Hz: $t(10) = 2.73$, $p < .05$] although those for the synchronization were not [262 Hz with 393 Hz: $t(10) = -1.34$, $p > .05$, 262 Hz with 524 Hz: $t(10) = 0.266$, $p > .05$] (Fig.2, right).

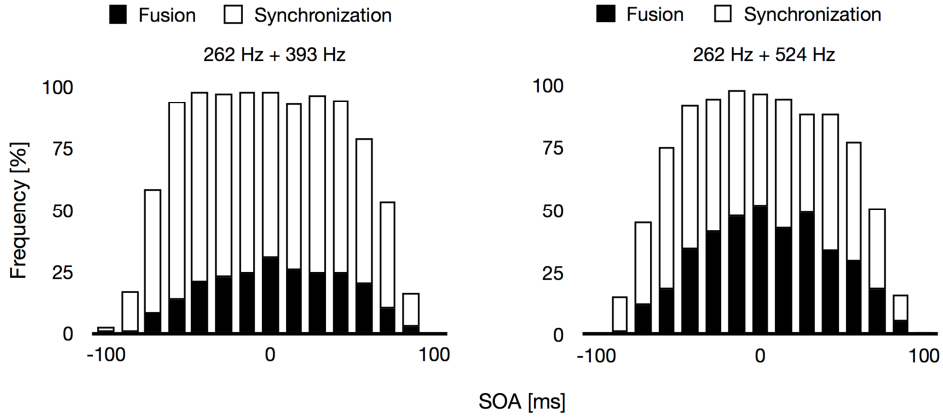


Figure 1: Frequency distribution of the fusional and synchronizing judgment. Each bar shows the proportion of each judgment against the SOA conditions. The negative value of SOA means that the higher frequency tone precedes the lower frequency tone, and positive value of SOA means that the lower frequency tone precedes the higher frequency tone. Left panel shows the results of the condition of 262 Hz and 393 Hz tones. Right panel shows the results of the condition of 262 Hz and 524 Hz tones.

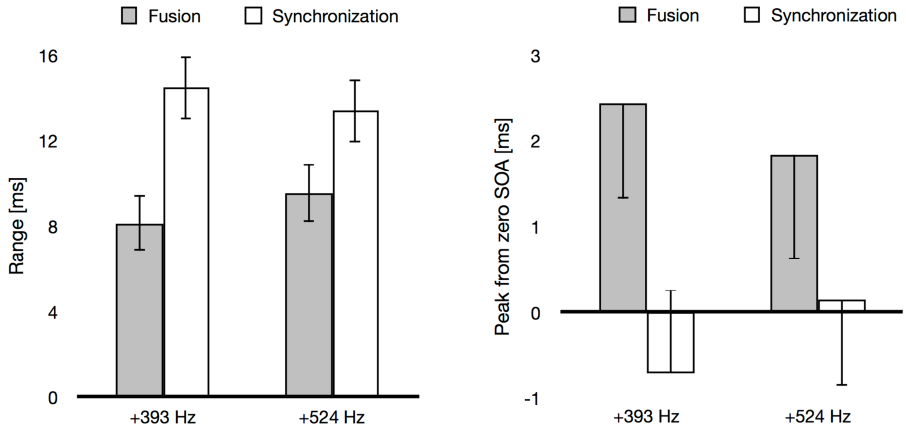


Figure 2: Left panel shows the range of fusional and synchronizing judgments. The ranges are obtained from the standard deviations of the frequency distribution for each judgment. Left side of the panel shows the results for the condition of 262 Hz with 393 Hz tones, while the right side of the panel shows the results for the condition of 262 Hz with 524 Hz tones. The error bar shows the standard error of mean. Right panel shows the peak of fusional and synchronizing judgments from zero SOA. The left side of the panel shows the results for the condition of 262 Hz with 393 Hz tones and the right side shows the results for the condition of 262 Hz with 524 Hz tones. The error bar shows the 95 % confidential interval.

4. DISCUSSION

We confirmed that simultaneity perception consisted of two processes, that is, the processes of fusion and synchronization. Both 262 Hz with 393 Hz and 262 Hz with 524 Hz tones conditions showed that both the fusional judgment and synchronizing judgment exist in these conditions. The fusional judgments increased with the decrement of SOA between the tones. This result indicates that listeners can perform the simultaneity judgment task by the use of the fusional process only, and that the synchronizing process may not be involved in the simultaneity judgment task.

The present study showed that the range of the synchronizing judgment was wider than that of the fusional judgment. As described in introduction, the simultaneity ranges reported by Parker (1988) were wider than the range reported by Wojtczak et al. (2012). Parker (1988) used the frequency sets with the various frequency intervals although Wojtczak et al. (2012) used that with the simple integer ratio. In addition, Wojtczak et al. (2012) described, “Listeners were encouraged to use whatever cue they found most effective for judging the synchrony between the tones”. These implies that participants in the experiments of Parker (1988) might judge the simultaneity by the use of the synchronizing process while participants in the experiments of Wojtczak et al. (2012) might judge the simultaneity by the use of the fusional process.

Our results showed that frequency separation between the tones did not affect the range of the fusional judgment and synchronizing judgment. This result is compatible with the results of Wojtczak et al. (2012). However, the results of Parker (1988) showed that the frequency separation between the tones affect the range of simultaneous judgment. As we showed, listeners can use both the fusional cue and synchronizing cue for performing simultaneity judgment task. If the decrement of frequency separation increases the weight of fusional cue, it is possible that listeners tend to judge the simultaneity for the two tones by the use of the fusional process. Furthermore, we showed that the range of the fusional judgment was narrower than that of the synchronizing judgment. Taken together these points, the effect of frequency separation on the range of simultaneous judgment reported by Parker (1988) could be obtained as a consequence of weight change for the fusional and synchronizing processes.

Our results showed that the peak SOA of fusional judgment is separated from zero SOA although this was not true for the synchronizing judgment. These results may be explained by the existence of the cochlea delay effect. It is well known that cochlea delay desynchronizes two

tones that differ in frequency. If this desynchronization affects the perception of relative timing for two tones, the peak SOA of fusional judgment and synchronizing judgment must be apart from zero SOA towards positive SOA because the lower frequency tone is physically delayed relative to the higher frequency tone at the cochlea processing. Woytczak et al. (2012) suggested that cochlea delay is compensated by higher stage of neural processing and does not affect the perception of simultaneity. Our results that the peak SOA was separated from zero SOA only for the fusional judgment suggest that cochlea delay is compensated between the fusional processing and synchronizing processing.

Both the temporal order process and simultaneity process for two tones concern with the relative timing between the tones. Mossbridge, Fitzgerald, O'Connor, and Wright (2006) showed that the mechanism which underlies the temporal order perception is different from that for the simultaneity perception. Using a perceptual-learning paradigm, they trained their participants to improve accuracy of temporal order / simultaneity judgment and test whether this improvement were generalized to other task. Their results implied that improving performance for a specific task would not facilitate the performance for the other task. However, two-category simultaneity judgment task they used are not adequate to reflect listener's perceived change because simultaneity perception should be accounted by three categories. In addition, Hirsh (1974) suggested that temporal order perception for two tones was not based upon a single process. Then, it is still probable that temporal order process and simultaneity process have some common basis.

Further work is required to understand how frequency separation affects the perception of simultaneity. Varying frequency separation between tones would also vary the frequency ratio, tonal consonance, musical consonance, and harmonic relation. In order to explore the relationship between simultaneity perception and frequency separation, it is necessary to clarify the quantitative relation between these possible factors and the processes of fusion and synchronization.

5. CONCLUSION

We investigated the mechanism for the perception for simultaneity by the use of three categorical judgments (fusion, synchronization and asynchrony). We showed that the perception of simultaneity consists of the processes of perceptual fusion and perceptual synchronization. The range of the fusional judgment was narrower than that of the synchronizing judgment. The frequency separation did not affect the range of each judgment but the peak. Our results suggest that the desynchronization caused by cochlea delay was still observed for the fusional processing. Our results also suggest that cochlea delay did not affect the synchronization process because compensation process should exist after the fusional processing, and before the synchronization process.

6. REFERENCES

Hirsh, I. J. (1974). Temporal order and auditory perception. In Sensation and measurement *(pp. 251-258). Springer Netherlands.*

Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., & Wright, B. A. (2006). Perceptual-learning evidence for separate processing of asynchrony and order tasks. The Journal of neuroscience, 26*(49), 12708-12716.*

Noorden, L. V. (1975). Temporal coherence in the perception of tone sequences *(Doctoral dissertation, Voorschoten:[sn]).*

Parker, E. M. (1988). Auditory constraints on the perception of voice onset time: The influence of lower tone frequency on judgments of tone onset simultaneity. The Journal of the Acoustical Society of America, 83*(4), 1597-1607.*

Sinico, M. (2004). Influence of bichord interval on perception of simultaneity. Perceptual and motor skills, 99*(3), 937-944.*

Wojtczak, M., Beim, J. A., Micheyl, C., & Oxenham, A. J. (2012). Perception of across-frequency asynchrony and the role of cochlear delaysa). The Journal of the Acoustical Society of America, 131*(1), 363-377.*

[1-04] THE EFFECT OF AUDIO PANNING ON THE PERCEPTION OF MOVEMENT IN FILM

Hyun Sung NAM[1]
[1]*City University of New York, USA*

Poster

Sound and vision are closely interrelated, affecting each other; to complement, and at times to distract. It is an important phenomenon to study, especially regarding art forms that involve multiple senses; like film. Sound itself is capable of enhancing the sense of space and movement with manipulations in reverberation, volume, timbre, synchronization, and panning. This study was specifically motivated by the concept of audio panning automation synchronized with visual movements. In other words, sound "moving" while the image moves. The goal of this research was to find the optimum degree of audio panning through subjective testing. 18 subjects participated and among them were 9 students studying music and/or sound technology. The subjects were asked to watch short movie clips in pairs and answer a survey asking "Which of the two sound more natural".

The experiment in whole was divided in to two large sections, each section using one movie clip as the stimulus. Each clip shows auto vehicles moving horizontally from one end of the screen to another without stopping. Three different audio panning degrees were implemented (30,75,100), rendered into the video, and paired randomly for the subjects to view and assess. Section 1 used an excerpt from the feature film "Hurt Locker" featuring an army truck moving at 20miles/hour, and section 2 used an exempt from a short film "Monticello Track Attack" featuring a race car moving at 150 miles/hour. In both cases, among the three panning degrees;narrower than the screen(30 deg), slightly wider than the screen (75 deg), wider than the screen(100 deg), 75 and 100 panning degrees were preferred over 30, but between 75 and 100, the difference was not as significant as when compared with the narrowest 30 degrees. Within the subject pool, the music/sound specialists gave more significant results while the non-musician pool answered rather randomly. Also, section 1 results showed more significance compared to section 2 due to the speed difference between the two stimuli.

[1-05] AMBIGUOUS

Iku Nemoto[1]*, Satoshi Ueno*[1]
[1]*Tokyo Denki University, Japan*

Poster

The purpose of the present work is to show the ambiguity between two imaginary melody lines and their perception observed by behavioral and neurophysiological experiments. Two sustained notes, A4 and E5, and a repetition of C#5 given by eighth notes alternating with eighth rests, were proposed to make up ambiguous music; one often hears an imaginary melody of repetition of E5-C#5 or of A4-C#5 although the sustained

notes (A4 and E5) are never repeated. Sometimes one hears only the repeated C#5 just as actually presented. MEG (magnetoencephalography) measurements during presentation of this ambiguous music showed different (although not always significantly different) responses to the last note of the stimulus depending on whether or not the last note was congruent with the imaginary melody which the subject was induced to perceive. Behavioral experiments yielded the transition time among the three inner states of the subject corresponding to the three phrases the subject heard. The obtained transition time distribution was used to estimate the state transition probability based on a simple model. This kind of ambiguity between imaginary melodies has never been proposed at least to the authors' knowledge and may provide new clues or methods for investigations of melody perception.

[1-06] COMPILING AND PROCESSING THE YALE-CLASSICAL ARCHIVES CORPUS

Christopher White[1], Ian Quinn[2]
[1]University of North Carolina, Greensboro, USA, [2]Yale University, USA
Poster

This presentation will introduce the Yale-Classical Archives Corpus (YCAC), a collection of data for research on harmony and rhythm in Western classical music. This corpus is based on data from classicalarchives.com, a repository of thousands of user-generated MIDI files of pieces from all periods of Western music history. The original data is proprietary and remains on the classicalarchives.com repository. YCAC makes available metadata for each MIDI file, as well as a list of pitch simultaneities ("salami slices") in the MIDI file. Metadata include the piece's composer, date of composition, genre (e.g., symphony, piano sonata, nocturne, etc.), the composer's country of origin, instrumentation, meter, and key. The processing step groups the file's pitches into vertical slices each time a pitch is added or subtracted from the texture, recording the slice's highest pitch, lowest pitch, prime form, scale-degrees in relation to the global key, and the event's offset (measured in the number of quarter notes separating the event from the file's beginning). The pitch content of each slice is recorded in a format readable by the music21 Python library (a library designed for music processing and data mining). The major benefit of the YCAC is its sheer size. The corpus contains 13,769 MIDI files by 571 composers yielding over 14,051,144 vertical slices. While there are several inconsistencies associated with compiling user-generated or open-source material (e.g., a large amount of encoding errors and incorrect metric information) we propose error-detection and meter-finding algorithms to mitigate these problems. For instance, by tracking the probability distributions of prime forms, pitches, and scale degrees, we can identify corrupted data by flagging files (or parts of files) whose distributions significantly differ from the remainder of the corpus (essentially treating encoding anomalies as noise within an otherwise statistically consistent signal). We also use a meter-finding algorithm that identifies the periodicities with which simultaneous note onsets recur, thereby identifying the prevalent meter and potentially identifying metrical changes within a file.

[1-07] EXAMINING THE ANTICIPATORY: META-ANALYSIS OF RESPONSE TO STRUCTURAL FUNCTIONS

Dale Misenhelter[1]
[1]University of Arkansas, USA
Poster

This meta-analysis explores findings from previous and current preference and response studies by the author, with a particular interest in examining micro-data (specific time segments) within longer response patterns that utilize(d) overt melodic or harmonic content as a means of signaling approaching structural functions. Listener anticipation and response is assumed to take shape as musical events occur in real time. Data collection in the studies utilized traditional musical works identified as having strong thematic (melodic) sequences, primarily (though not exclusively) in recapitulation or coda structure. Harmonic function at similar points in time eliciting tension and anticipation were also examined through the lens of continuous response data and graphic analyses.

Variables considered in the primary research question also include 1) familiarity of participants with select musical excerpts (via comparative Likert-type scale self-report data) and the possibility of comparative cognitive representations; 2) the general experience level of participants (often characterized as musicians and non-musicians) as a consideration; and 3) overt musical elements presented (melodic prompts, musical event frequency, tension, textural and dynamic considerations, harmonic functions, etc). Specific works analyzed included the Bach Passacaglia in C minor, Beethoven Symphony 7 (2), Mussorgsky Pictures at an Exhibition, and Stravinsky Firebird Suite.

A secondary research question was in regard to focus of attention - the ability of listeners to (continue to) respond over time to elemental variables. The ability of listeners to distinguish between perceived musical elements or other stimuli while concurrently attending and responding is a process loosely termed "multi-tasking." While there is considerable research on listener ability to discriminate and/or prioritize among elements in aural environments, introducing anticipation as a measurable component of listener response and possibly distinct from tension is evidenced comparatively little in the response research literature.

[1-08] EMBODYING THEORETICAL RESEARCH IN MUSIC COGNITION: FOUR PROPOSALS FOR THEORY-DRIVEN EXPERIMENTATION

Andreu Ballús[1], Eric Arnau[1,5], Oriol Nieto[3], Frederic Font[2], Alba G. Torrents[1,4]
[1]Universitat Autònoma De Barcelona, Spain, [2]Pompeu Fabra University, Spain, [3]New York University, USA, [4]Universidad Nacional De Córdoba, Argentina, [5]York University, Canada
Poster

ABSTRACT

Research in the field of music cognition typically focuses either on low-level, technically oriented approaches or on highly abstract ontological discussions that lack direct grounding in evidence. To bridge this gap, we propose a revision of the ontology underlying such research, from a perspective restricted to the acoustic and individual aspects of music to an embodied, extended, and anti-individualist approach. We explore the application of these ideas to empirical research in a twofold way: by discussing two experiments conducted by our group and by proposing two ideas for further experimentation. One of the conducted experiments tests whether the ability to play an instrument in any of its dimensions has

an influence on how a subject listens to music; the other one explores the impact of visual information on the perception of sound as music. We comment on the results obtained and their theoretical significance. Our work shows that it is possible for abstract theorizing and concrete experimentation to go hand in hand in the field of music studies.

1. INTRODUCTION

Research in the field of music cognition typically consists of low level, technically oriented approaches focused on very specific aspects of musical structure and/or of psychoacoustic processes (e.g.: Crummer et al., 1988; Bigand et al., 2001; Demorest & Morrison, 2003; Janata, 2009), and, on the other end of the spectrum, of highly abstract ontological discussions (e. g.: Hegarty, 2001; Van Nort, 2006). Moreover, due to the relative disconnect between these lines of inquiry, most technically work tends to uncritically assume a particular ontology of music, according to which it would be a fundamentally acoustic phenomenon (Mannes, 2011; Downie et al., 2009).

While acknowledging the value of these approaches and of other narrowly framed contributions (such as the growing body of technical work on music information retrieval), we believe that the field of music cognition benefits greatly, at the present stage, from those studies in which abstract theory and specific data are put to work together, and that it is important to keep checking our implicit high-level theoretical models against empirical reality. In order to do so, our work focuses on reviewing some of the implicit assumptions of existing research in music cognition and elaborating a contrasting set of thesis whose heuristic and explanative capacities could be tested through explorative research. These theses draw heavily on recent developments in what has come to be called the third generation of cognitive science, including philosophical and theoretical positions such as distributed cognition (Hutchins, 1995), the embodied mind thesis (Varela, Thompson & Rosch, 1991), the extended mind thesis (Menary, 2010) and enactivism (Thompson, 2007). Some aspects of these developments have already been applied to music research in recent years (Schellenberg and Trehub, 1999; Cano, 2006; Leman, 2007; Clarke & Clarke, 2011). However, most of these applications focus on specific theses and do not present a general alternative to the core tenets shared by the classical cognitivist stance and its connectionist modification, both of which present musical cognition as a matter of individual symbol processing. Our aim is instead to maintain a broader perspective, using explorative approaches not to reach a definitive conclusion on any technical detail, but to test how does our alternative set of suppositions stand when confronted with empirical data (compared to more classical approaches).

The core of our alternative high-level model is the thesis that music is not a strictly acoustic or psycho-acoustic phenomenon in any restrictive sense (a good initial discussion along these lines can be found in (Wiggins, 2009)). Some of its crucial aspects have nothing to do with either wave propagation or the individual psychological perception of music. This unfolds into three sub-theses: (1) Details of physical embodiment traditionally considered irrelevant to cognitive phenomena are relevant to music cognition (*Embodied Music Thesis*); (2) Interaction with technical instruments and with the environment is relevant to musical phenomena in a way that goes beyond merely enabling the production and transmission of certain sounds (*Embedded / Extended Music Thesis*); (3) Some forms of interpersonal interaction not directly related to the production and propagation of sound are relevant to musical phenomena (*Musical Interactionism / Anti-individualism).* The experiments presented in this work are designed to explore musical cognition phenomena under this framework.

2. EXPERIMENT A: THE EFFECT OF MUSICAL PRACTICE ON MUSIC PERCEPTION

Our first experiment explores the impact of musical practice to musical perception, with the aim of assessing whether the underlying relevant factors have to do with the bodily engagement with artifacts and its coordination with others.

2.1. Methodology

The experiment was conducted online, with a sample of 110 subjects of ages ranging 18 to 63 from different nationalities. The first section consisted of a survey to profile the subjects' relation with music, both as performers and as listeners. To avoid falling back into the assumptions of the cognitive/acoustic model, we considered a wide range of dimensions of musical knowledge, production and reception while designing the test, including different levels of bodily implication, degrees of intentionality and individual and collective aspects. To establish their performer profile we had subjects report whether they have musical studies and whether they perform music of any sort, whether they play any instrument, they sing, they perform in public, and they perform with other people. To establish the profile of listening practices, we had subjects report the amount of weekly hours that they listen to music, and the proportions of that listening time that is devoted to listening selected/non-selected music, and to listening music with full/peripheral attention.

In order to analyze how participants perceive music, subjects were asked to listen to several audio samples of diverse musical genres. Subjects had to (i) report the number of voices they could identify in a given sample; (ii) report the number of segments in which a given sample could be segmented; (iii) provide a series of tags describing several attributes of the sample songs; and (iv) provide a number of tags identifying both the similarities and the differences between two given sample songs. These responses were considered in relation with the *know-how* and *know-to* related variables, using supplementary information as educational level and listening practices as a contrast

2.2. Analysis of Results

2.2.1. Identifying Voices and Segments

We first investigated which dimensions of music performance have effects on the discrimination of different voices in a sample with a series of nonparametric tests. Wilcoxon Rank Sum test shows that performing music in any form has a significant effect (assuming $\alpha = .05$) on the number of identified voices (Nyes = 56, Nno = 45, $W = 950$, p = .033), and so does having music studies (Nyes = 59, Nno= 42, W = 824.5, p = .004). However, the effect doesn't seem to be large, considering the differences shown in figure 1.

When decomposing the notion of performing music into further dimensions, we found that playing an instrument ($N_{yes} = 56$, $N_{no} = 45$, $W = 950$, $p = .034$) and singing ($N_{yes} = 43$ $N_{no} = 58$, $W = 860$, $p = .007$) also have a significant yet seemingly small effect on voice discrimination, as illustrated in figure 2. We also found that performing in public has a significant effect on voice discrimination ($N_{yes} = 31$, $N_{no} = 24$, $W = 221$, $p = .010$), while performing with other people does not ($N_{yes} = 35$, $N_{no} = 18$, $W = 250$, $p = .223$). Figure 3 illustrates these results.

We performed two-way ANOVAs to further assess the independence of these factors. We found a significant effect of having music studies ($F(1,149) = 5.30$, $p = .023$) as opposed to performing (ns), and a significant effect of performing in public ($F(1,49) = 4.26$, $p = .044$) as opposed to playing with others (ns). No significant interactions between the factors were found. Playing an instrument and singing showed no significant effects. Figure 4 illustrates the pairs where significant effects were found.

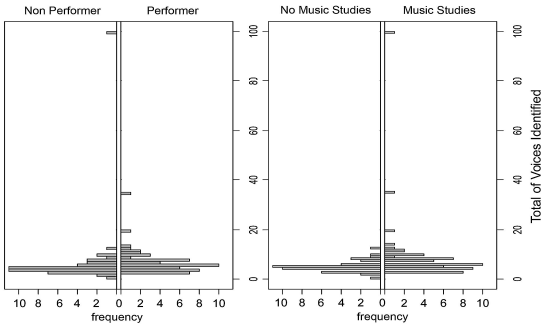


Figure 1: Paired histograms showing the number of voices identified by Performers/Non Performers (left) and subjects with Musical Studies and without (right).

None of the factors involving music production showed significant effects on the discrimination of segments on a different sample. Finally, no factors involving the listening practices of the subjects (total amount of listening hours and relative amounts with different degrees of control or attention), showed any significant effects on the discrimination of voices or segments.

These results show that, while people's habits and practices of listening to music have no significant relation with their capacity to discriminate voices, the latter is related to whether they have musical studies and to whether they perform music. While the extent of such effects remains uncertain, our results seem to suggest that the know-what aspect that comes with music studies is a more relevant factor than the embodied and know-how aspect that comes with musical practice.

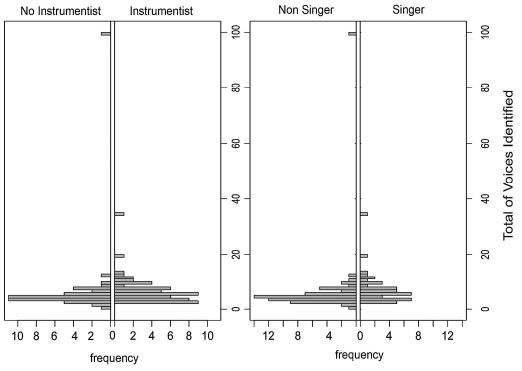


Figure 2: Paired histograms showing the number of voices identified by Instrument player / Non Instrument player (left) and Singers/Non Singers (right).

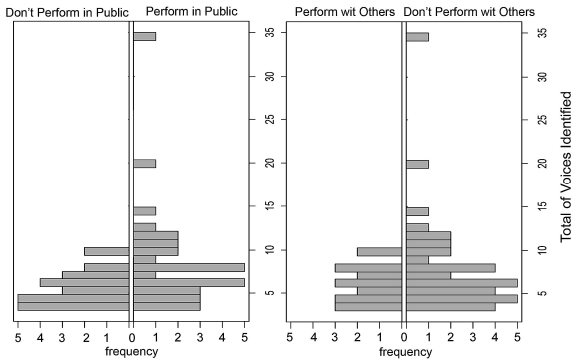


Figure 3: Paired histograms showing the number of voices identified by subjects who perform / do not (left) and subjects who play with others / do not (right).

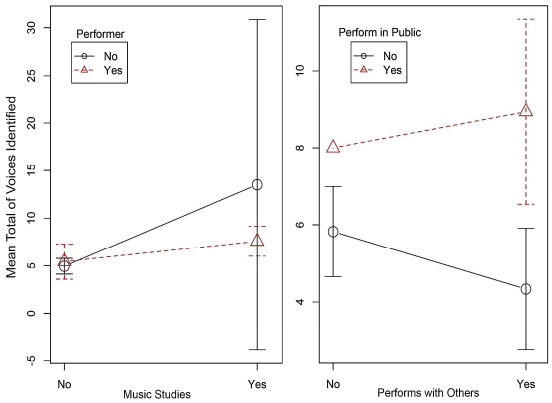


Figure 4: Mean number of voices identified by subjects decomposed according to the 'Performer' and the 'Musical Studies' partitions of the sample (left) and to the 'Performance with Others' and the 'Public Performance' (right).

2.2.2. Describing and Comparing Tracks Using Tags

In order to assess what features subjects perceive as more salient when listening to music, the responses obtained on the open descriptive and comparative questions were categorized into different semantic fields as depicted in figure 5.

We also partitioned all tags into two second order categories: 'Acoustic' encompasses tags referring to features that can allegedly be reduced to acoustic or psychoacoustic properties –those in falling within the categories of Rhythm, Timbre, Structure and Musical Attributes (melody, pitch, etc.)–, and 'Non-Acoustic' includes the remaining tags that cannot refer to features heavily relying on extra-acoustic features. Overall, as figure 6 shows, most judgments involve non-acoustic features.

With regard to the amount and variety of the responses, our first finding is that the total number of tags is significantly related to performing music (N_{yes} = 43, N_{no} = 58, W = 936, p = .032), but not to having musical studies. More specifically, nonparametric tests indicate that playing and instrument does not show significant effects, but singing does (N_{yes} = 54, N_{no} = 47, W = 938, p = .024). Moreover, performing a two-way ANOVA for the factors Instrument and Singer with regard to the total of tags, shows that Instrument is a significant factor (F(1, 97) = 4.73, p = .032), and being a singer is not (F(1, 97) = 0.00, p = .959), but there is a significant interaction effect (F(1, 97) = 4.27, p = .041).

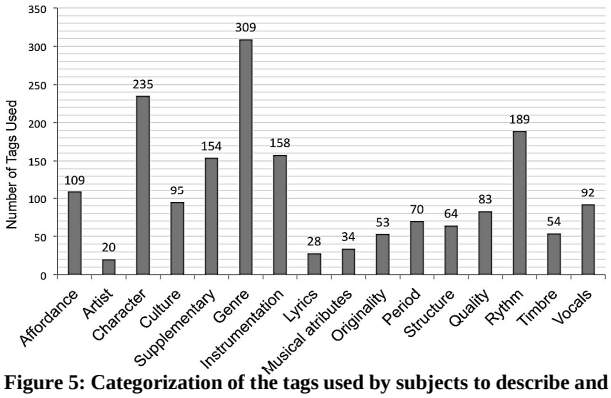


Figure 5: Categorization of the tags used by subjects to describe and compare music samples, with their frequencies. We take the categories to be self-explanatory, except for 'Supplementary', which stands for supplementary, non-conventional symbolic or metaphoric projections.

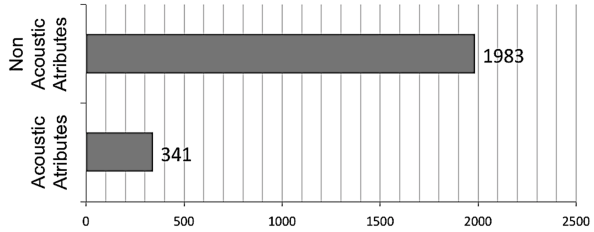


Figure 6: Second order categorization of the tags used by subjects, with their frequencies.

If we analyze the amount of different categories used by subjects instead of the number of tags, the results are less clear. It appears that the only factor close to statistical significance is playing an instrument (N_{yes} = 56, N_{no} = 45, W = 973.5, p = .049). However, the results of the ANOVA take out its significance (F(1, 97) = 3.41, p = .068). Neither performing in public or with others, or any of the variables related to listening practices showed any effect on the overall prolificacy and semantic diversity of subjects' responses.

So, there is not strong evidence about that people who make music have a more prolific and diverse impression of music. Nevertheless, more interesting results come up when we target specific categories of musical judgment. We found effects on the usage of tags related to 'Genre', 'Instrumentation', 'Musical Attributes' and 'Quality'. Music performers significantly more often choose these qualities.

Using Fisher's Exact Test (two-sided) we found that musical practice has a significant effect on the frequency of genre ascriptions when describing and comparing samples (p = .046), while having music studies does not show such effect. On a finer level, playing an instrument has a significant effect (p = .024), while singing does not. Such effects are confirmed by the ANOVA, which shows 'Performer' to be a significant factor (F(1, 97) = 8.87, p = .003) as opposed to 'Music Studies' (ns) , and 'Instrument' to be a significant factor (F(1, 97) = 8.39, p = .005) as opposed to 'Singer' (ns). No interaction effects were found in either case (figure 7).

The same pattern appears if we consider the amount of judgments referring to musical attributes (melody, pitch etc.). Performing music has a significant effect (p = .016), while having music studies does not. On a finer level, playing an instrument has a significant effect (p = .002), while singing does not. Such effects are again confirmed by an ANOVA, which shows 'Performer' to be a significant factor (F(1, 97) = 9.58, p = .003) as opposed to 'Music Studies' (ns), and playing an instrument to be a significant factor (F(1, 97) = 9.55, p = .003) in front of 'Singing' (ns). No interaction effects were found in either case (figure 8).

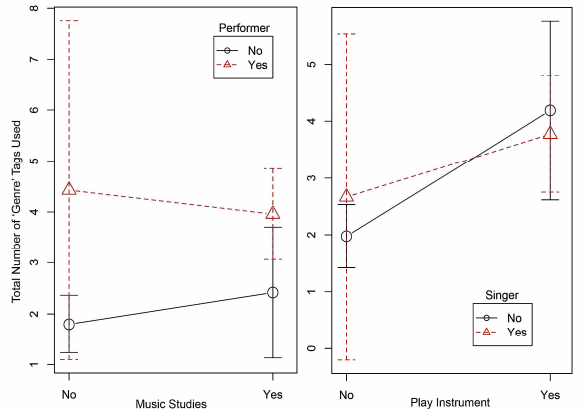


Figure 7: Mean number of 'Genre' tags used by subjects decomposed according to the 'Performer' / 'Musical Studies' partitions of the sample (left), and to the 'Instrument' / 'Singer' (right).

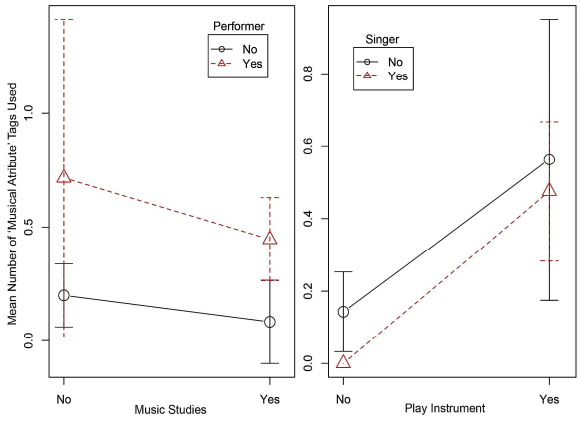


Figure 8: Mean number of 'Musical Attribute' tags used by subjects decomposed according to the 'Performer' / 'Musical Studies' partitions of the sample (left), and to the 'Instrument' / 'Singer' (right).

In contrast with this pattern, different factors appear to be relevant for the production of judgments involving instrumentation. Here we find that both being a performer (p = .004), and having music studies (p = .003) have significant effects. However, the ANOVA between these two factors indicates that only the former effect is significant (F(1, 49) = 11.13, p = .001). Unlike the previous cases, when we decompose 'Musical Performing' into playing an instrument and singing, the categorical variable (i.e. whether subjects make judgments about 'Instrumentation') shows a significant relation with singing (p = .008), but not with playing. Yet the ANOVA points to the opposite direction, indicating that playing an instrument is the statistically significant factor (F(1, 97) = 4.86, p = .030). No interaction effects were found in either case. Moreover, no effects were found on performing in public, performing with others, or having different listening habits. Similar results are found with regard to the production of quality judgments. Both being a performer (p = .026), and having music studies (p = .004) have significant effects.

As a result, what we notice first is that making music is strongly coordinated with the type of features a subject focuses on. More interestingly, we see that the features that attract a special attention to instrument players –genre, melody– are different from those features that attract singers –instrumentation, quality. Not only playing an instrument and singing have different effects, but also the effects of singing seem to align with those of having musical studies. This might indicate that the effect of singing has more to do with knowledge and familiarity, and less with the embodied engagement with musical artifacts.

3. EXPERIMENT B: EFFECT OF VISUAL CUES ON MUSIC IDENTIFICATION

Experiment B focuses on the conscious identification of an acoustic stream as music, and tries to determine whether the perception of visual images with music-related content has any significant effect on it. Our aim is to heuristically test the consideration of non-acoustic, extended aspects of music as proper elements of it.

3.1. Methodology

A four minute custom piece of audio was created for experiment B, in which different musical elements slowly and subtly appear against a backdrop of noise, so that it sounds like noise at the beginning, and at some point it starts sounding musical. The same subjects (n=110) that performed experiment A were randomly assigned to one of two conditions. Subjects in condition A were asked to listen to the piece of audio and report the exact moment in which they started to perceive music. Subjects in condition B were asked to watch a video clip with the same audio paired with footage from an experimental music performance, and to report, equally, the moment in which they started to perceive music29.

3.2. Analysis of Results

Probability density functions (PDF) were calculated for the music detection time reported by both groups. PDF was estimated using Kernel Density Estimation. Figure 9 depicts PDF(t) for the two groups. It shows three main t instants where participants tend to indicate that the audio track transitions from non-musical noise to music (t_1 ~30, t_2 ~100 and t_3 ~180). Those instants correspond to well-defined changes in the audio track, but no particular events occur in the video. In t_1 a slightly dissonant harmonic element with a low volume is introduced in the track. In t_2 a musical pad appears with a rather strong presence in the mix. Finally, at t_3 the background noise disappears from the audio track and only a melody is left. Both PDFs have their maximum values around t_1, but PDF of group B has stronger peaks in t_2 and t_3 than PDF of group A. On average, participants listening to the audio track without visual information tend to indicate that transition from non-musical noise to music occurs 18 seconds before than other participants, but the comparison is not statistically significant (F= 320, p = .650, RankSum Test). Overall these results suggest that the visual information may be acting as a distracting factor and thus some participants of group B might not be noticing the change in t_1.

To get more insight on this analysis, we further divided groups A and B according to their musical studies (yes/no). Figure 10 shows the estimated PDF(t) for the four resulting groups. As it can be seen, participants with musical studies (regardless of the group) also have a tendency to indicate that transition to music happens earlier than participants without musical studies. In fact, the comparison happens to be stronger than when comparing groups A and B. On average, participants with musical studies indicate that transition from non-musical noise to music occurs 36 seconds earlier than participants without musical studies (F = 160.5, p = .006). This result indicates that having musical studies has a bigger

29 URL to the video clip: https://www.youtube.com/watch?v=dSirlyqLR40

impact than being exposed to intentional visual information paired with the audio. One possible explanation is that the slightly dissonant harmonic element introduced at instant t_1 on the audio track is probably not being considered as music by participants without musical studies.

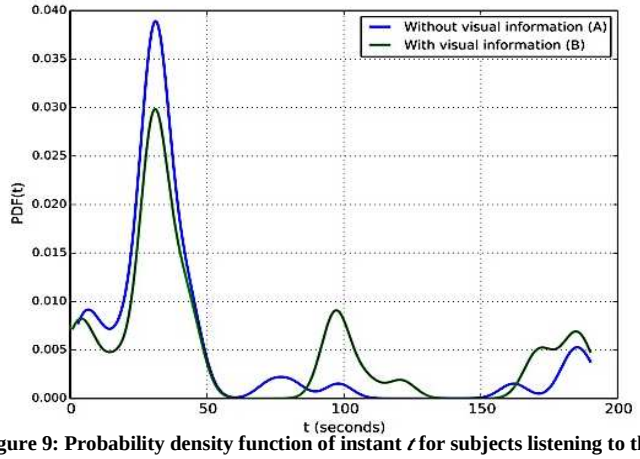


Figure 9: Probability density function of instant t for subjects listening to the sample audio with and without visual information.

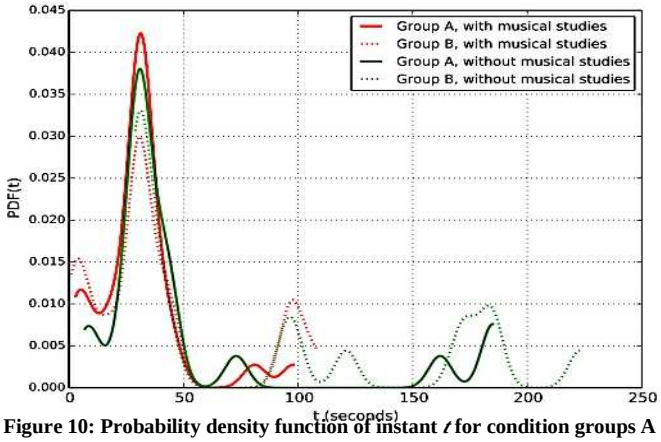


Figure 10: Probability density function of instant t for condition groups A and B decomposed into those with music studies and those without.

4. EXPERIMENTS DISCUSSION

We take our findings to support our contention that research in musical cognition can be illuminated by adopting a framework that distances from the common assumption that music is a fundamentally an acoustic phenomenon. First, they show that music practice has several effects on musical perception. People who make music seem to have a different capacity to discriminate voices, and they perceive some features as more salient. This is still more relevant against the fact that none of these effects were found for different listening practices. It seems that listening to more or less music, or in different conditions, does not affect these results. We also found that having music studies has several significant effects, but they do not align with the effects of music performance. This discrepancy between the effects of the domain of the abstract know-what and the domain of the embodied know-how can be accounted for in terms of our *embodied music thesis* –details of physical embodiment traditionally considered irrelevant to cognitive phenomena are relevant to music cognition.

However, the dimension of making music is still too broad. By considering further dimensions of musical performance, we see that across our different tests, the effects exhibited by playing an instrument differ quite systematically from those observed for being a singer, which align much more with the effects found by having musical studies. We argue that this pattern points to the fundamental distinction between the bodily engagement with an external artifact that comes with playing an instrument, and the practice of singing, in which there is no integration of such external scaffolds. This goes along the lines of our embedded/extended music thesis: interaction with artifacts and environmental scaffolds are relevant to musical phenomena in a way that goes beyond merely enabling the production and transmission of certain sounds.

In order to assess our musical anti-individualist thesis –some forms of interpersonal interaction not directly related to the production and propagation of sound are relevant to musical phenomena–, we considered two further variables: whether subjects perform in public and whether they play with others. With that regard, only a barely significant effect was found that might relate public performance with a tendency to focus more on acoustic features of music. No effects were observed from playing with others. But our impression is that, unlike the embodied and extended dimensions of musical practice, which have a more longstanding transformative effect, the collective dimension of music might well be a more fleeting phenomenon, which, ideally, should be observed on the fly. We next sketch some ways to circumvent this and other limitations of our current experiments.

5. CONCLUSIONS AND FURTHER RESEARCH

The experiments we performed seem to confirm the heuristic fruitfulness of our general theoretical model (i.e. that music is not a strictly acoustic or psycho-acoustic phenomenon), but some of the results are certainly not informative enough. To correct this, and as an indication of possible directions for further research, we propose two ideas for follow-up experiments.

Proposed experiment C focuses on the relationship between music discrimination and the subject's listening practices while making use of the work already done on the analysis of music descriptions and categorizations: in a context and setting similar to that of the performed experiments, subjects are provided with pairs of musical segments of different styles and with different relationships (same musical style, performances of the same piece, etc.). They are asked to provide detailed information on their listening practices and to answer short open questions on the paired segments. In one group the pieces are accompanied by some "exemplifying" descriptions; the influence of this elements in the responses and its interaction with listening practices is analyzed in contrast with the control group.

Proposed experiment D would be an extension of the experiment designed to explore the relationship between the perception of sound as music and the concurrent activity of the listeners. Different groups of listeners are asked to perform (i) a simple puzzle-solving task, (ii) a puzzle-solving-task involving sound elements and (iii), a puzzle solving-task in coordination with another subject, and the influence of these practices in the perception of sound as music is compared to that of the control group.

These, of course, are just some potential directions in which further research could be developed on the basis of the exploratory approach we have presented. We encourage researchers to follow any of these potential paths and to contribute their testing of the heuristic value of our basic alternative model.

6. ACKNOWLEDGMENTS

The authors would like to thank Isam Alegre and Obsidian Kingdom for the music track created specifically for our experiment.

7. REFERENCES

Aucouturier, J. (2007). Sounds like teen spirit: Computational insights into the grounding of everyday musical terms. In J. Minett & W. Wang (eds.). Language, Evolution and the Brain, *Honh Kong, City University of HK Press, 35–64.*

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. Cognition, *81(1), B11-B20.*

Crummer, G. C., Hantz, E., Chuang, S. W., Walton, J., & Frisina, R. D. (1988). Neural basis for music cognition: Initial experimental findings. Psychomusicology: A Journal of Research in Music Cognition, *7(2), 117.*

Clarke, D., & Clarke, E. (Eds.). (2011). Music and Consciousness: Philosophical, Psychological, and Cultural Perspectives. *Oxford University Press.*

Demorest, S. M., & Morrison, S. J. (2003). Exploring the influence of cultural familiarity and expertise on neurological responses to music. Annals of the New York Academy of Sciences, *999, 112-117.*

Downie, J. S., Byrd, D., & Crawford, T. (2009). Ten years of ISMIR: Reflections on challenges and opportunities. In Proceedings of the 10th International Conference on Music Information Retrieval *(pp. 13–18).*

Hegarty, P. (2001). Noise threshold: Merzbow and the end of natural sound. Organised Sound, *6(3), 193-200.*

Hutchins, E. (1995). Cognition In The Wild. Cambridge, MA: MIT Press.

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. Cerebral Cortex, *19, 2579-2594.*

Leman, M. (2007). Embodied music cognition and mediation technology. *Cambridge, MA: MIT Press.*

Mannes, E. (2011). The power of music: Pioneering discoveries in the new science of song. Bloomsbury Publishing USA.

Menary, R. (Ed.). (2010). The extended mind. *Cambridge, MA: MIT Press.*

Cano, R. L. (2006). 'What kind of affordances are musical affordances? A semiotic approach'. In: Simposio Internazionale sulle Scienze del Linguaggio Musicale. Bologna, *23, 25.*

Schellenberg, E. G., & Trehub, S. E. (1999). Culture-general and culture-specific factors in the discrimination of melodies. Journal of Experimental Child Psychology, *74, 107-127.*

Tillmann, B., & Bigand, E. (1998). Influence of global structure on musical target detection and recognition. International Journal of Psychology, *33, 107-122.*

Thompson, E. (2007). Mind in life: Biology, phenomenology, and the sciences of mind. *Cabridge, MA: Harvard University Press.*

Van Nort, D. (2006). Noise/music and representation systems. Organised Sound, *11(2), 173.*

Varela, F. J., Thompson, E. T., & Rosch, E. (1991). The embodied mind: Cognitive science and human experience. *Cambridge, MA: MIT Press.*

Wiggins, G. A. (2009). Semantic Gap?? Schemantic Schmap!! Methodological Considerations in the Scientific Study of Music. *In Proc of the 11th IEEE International Symposium on Multimedia, 477–482.*

[1-09] USE OF EIGENMUSIC TO ARRANGE MUSIC EXCERPTS

Mizuki Yamasaki[1], Masanobu Miura[2]
[1]Graduate School of Science and Technology, Ryukoku University, Japan, [2]Faculty of Science and Technology, Ryukoku University, Japan
Poster

ABSTRACT

When playing in an orchestral ensemble, musicians must arrange conventional, or large-scaled sheet music due to its inappropriateness in terms of musician assignment. Since the arrangement of sheet music requires certain musical knowledge and expertise, beginner musicians have difficulty in arranging sheet music appropriate for their needs. We propose an automatic arrangement method that involves using Eigenmusic, which contains eigenvectors for many music excerpts (*Abe et al*., 2012), this method is called "Abe's method". We extracted backing patterns from thousands of MIDI excerpts and constructed a database of the patterns to measure the similarity between phrases in current MIDI excerpts. The proposed method chops the tracks in the current MIDI excerpts into phrases, labels each phrase into melody, base, or backing, allocates musicians to each phrase, provides rests into appropriate positions on the phrases, and exchanges melodies among musicians to generate arranged

sheet music. Experimental results showed that the proposed method satisfied the naturalness of the music for listeners, but no significance of the proposed was confirmed for ease of playing.

1. INTRODUCTION

Orchestral performance involves four instrumental families; string, woodwind, brass, and percussion. It may contain approximately 50 to 80 musicians. Since a small number of musicians often play together as an ensemble, they need sheet music appropriate for their needs. In order to obtain a sheet music for limited number of players, arrangement is necessary. However, arrangement requires musical expertise and takes time to complete. Thus, automatic arrangement is preferred to solve this problem. Several studies of automatic arrangement have been reported, some of which are for arrangement based on melody extraction or instruments (*Shakoh et al*., 2010) and others are for arrangement of specific sheet music of wind orchestra into sheet music for a few musicians (*Maekawa et al*., 2006). However, previous studies did not focus on the naturalness of the arranged excerpts. Another method for automatic production of the base part of popular music from many MIDI excerpts by using principal component analysis (PCA) (*Abe et al*., 2012) was reported. Abe's method extracts tracks of the base guitar part from many MIDI excerpts and constructs a database of that part, which is used to extract the global average of patterns being used to generate the natural pattern of the base part. However, orchestral performances are not taken into account with this method.

Therefore, we propose an automatic arrangement method that takes into account the naturalness and easiness of playing arranged sheet music. In this study, naturalness of music is defined as the extent one can listen without feeling uncomfortable and easiness for musicians is defined as the how easy the music is to play.

Our research involved the following steps: i) database construction, ii) Eigenmusic calculation, iii) automatic arrangement, and iv) performance evaluation. For i), we constructed a database by using the extracted backing patterns from thousands of MIDI excerpts. For ii), we calculated the Eigenphrase of backing, which is a part of Eigenmusic, where the Eigenphrase of backing shows the average backing feature. For iii), we chopped the tracks of the MIDI excerpts into phrases on which we applied a previously proposed method (*Miura et al*., 2010) to obtain the phrases, this method is called "Miura's method". When labelling the phrases as melody, base, or backing, we used the Eigenphrase of backing. The labelled tracks are then composed into the arranged sheet music, where performance rests are allocated and musician exchanges are made. For iv), we confirmed the naturalness and easiness of the arranged music through an evaluation experiment. Figure 1 shows the process of the proposed method.

2. DATABASE CONSTRUCTION

A: Database construction

We constructed a music database for backing, called the "backing pattern database", where the included data are "backing pattern", which is the pattern between bars in the backing part. The data we used were from a set of 4147 MIDI excerpts of J-POP, and their onset, interval of note height, and MIDI-velocity were separately extracted in the database. The tracks in the MIDI excerpts were categorized into "Melody", "Base", "Backing", and "Perc.". Since the tracks of "Melody" and "Perc." are allocated in specific tracks in conventional MIDI format, they are not difficult to categorize. The "Base" tracks are automatically extracted with Abe's method, so the remaining tracks are categorized as a backing pattern.



Figure 1: Proposed method of automatic arrangement.

When obtaining the backing patterns, only 4/4 time is considered since 98% of J-POP MIDI excerpts are composed in 4/4 time (*Abe et al*., 2012). The minimum unit is sixteen notes in onset information. The shorter notes of length are rounded into the neighbouring unit of sixteen notes. The interval information sets up the first note of each bar as the standard and registers the difference in the remaining notes. The velocity information is obtained in the same steps of the interval information.

B: Method of calculating "Eigenphrase of backing"

"Eigenphrase of backing" is a coefficient matrix of principle components (PCs) calculated using PCA for the backing pattern database. The coefficient matrix of PCs is represented as a square matrix whose elements are $a_{p,n}$ and z_p and expressed as follows.

$$z_p = \sum_{i=1}^{n} a_{p,i} x_i , \qquad (1)$$

where R_p^2, a deviation of z_p, is represented as

$$R_p^2 = \frac{\sum_{i=1}^{n} \left(a_{p,i} x_i\right)^2}{n} . \qquad (2)$$

When R_p^2 is maximum or minimum, we can calculate $a_{p,n}$. We calculated the "Onset profile", "Interval profile", and "Dynamics profile" by obtaining the PCs for onset, interval, and velocity in the backing pattern database, respectively. The onset profile for the Eigenphrase of backing is that the row is the principal ingredient and the column is the onset time under the unit of sixteen notes. The interval profile for the Eigenphrase of backing and the dynamics profile for the Eigenphrase of backing are those whose rows and columns are principal ingredients and an indexes of notes, respectively. We considered the cases of 3 to 12 notes in one bar independently and analyzed their PCs independent of the number of notes. Finally, we obtained Eigenphrases as matrices of $\underset{16\times16}{O}$, $\underset{n\times n}{I}$, and $\underset{n\times n}{D}$, where n = 3, 4, …, 12 and O, I, and D represent the onset, interval, and dynamics profiles, respectively.

3. METHOD

The restrictions of the proposed method are described as follows. Melody and base must be monotone and played by only one player for each. More than two musicians can be assigned on the backing. We allocate rests for phrases to make the arranged sheet music easy to play. Specifically, we allocate rests in the phrases in the backing. In some cases, phrases are exchanged between musicians to implement a sort of dialogue of music, which is expected to improve naturalness.

A: Method of chopping tracks of current MIDI excerpt into phrases (*Miura et al*., 2010)

We chopped all the tracks in the current MIDI excerpt. It is based on Miura's method. Figure 2 outlines this method. Chopping is done for all tracks where the silent time is longer than the standard length. The method uses criteria on the basis of the Generative Theory of Total Music (GTTM) (*Nagashima et al*., 1999).

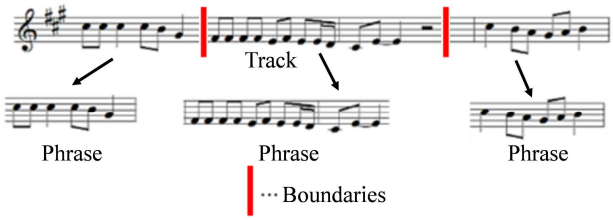


Figure 2: Outline of chopping phrases.

B: Categorizing phrases by Eigenphrase of backing

We label the phrases as melody, base, or backing by using the Eigenphrase of backing. First, we chop the track of the melody part into phrases by using Miura's method. After that, the boundaries of the chopped phrases are applied to the other tracks. Methods for calculating the extent of similarity and references needed for similarity calculation are required to label all the phrases. For the former, the Abe's method is used for calculating the similarity of phrases so that similarity calculation between phrases is possible. For the latter, we determine a set of standard tracks for melody, base, and backing. The standard melody track is simply the 1st track in MIDI format since the 1st track is defined as a melody track in conventional MIDI format. The standard base track is determined using the Abe's method, which robustly extracts base tracks from MIDI excerpts. The standard backing track is then determined by calculating the similarity among phrases. We obtain two similarities from the melody and base, and by observing the summation of the two similarities, we may find an isolated phrase that has the highest similarity score. Then the isolated phrase is used as the standard backing. We then obtain three standard tracks for melody, base, and backing. Finally, other tracks besides the three standards are labelled as melody, base, or backing based on the similarity scores.

We now explain the calculating method of the similarity scale. We first extract the onset, interval, and velocity within the phrases. We then obtain three vectors for onset (v_o), interval (v_i) and dynamics (v_d). By evaluating $v_o O$, $v_i I$, and $v_d D$, we obtain the PC scores, whose elements are $X_{t,s,n}$, and expressed as follows.

$$\left(X_{t,s,1}, X_{t,s,2}, \cdots, X_{t,s,n}\right) = \left(y_1, y_2, \cdots, y_n\right) \times \begin{pmatrix} a_{1,1} & \cdots & a_{n,1} \\ \vdots & \ddots & \vdots \\ a_{1,n} & \cdots & a_{n,n} \end{pmatrix}, \qquad (3)$$

where n is the number of elements within a bar, t is the bar ID, s is the track ID, and y_n is n elements of either onset, interval, or dynamics. We then calculate the Euclid distance of PC scores between phrases, labelled as $E_{t,s,k}$, and expressed as follows.

$$E_{t,s,k} = \sqrt{\sum_{a=1}^{n} \left(X_{t,s,n} - X_{t,k,n}\right)^2}, \qquad (4)$$

where *s* and *k* are the track IDs and *n* is the number of elements in the PCs. The smaller the $E_{t,s,k}$, the more *s* and *k* are similar. The $E_{t,s,k}$ is obtained for the three profiles then the average of similarity scores among the three profiles are calculated. We refer to the average when labelling the phrases as melody, base, or backing. The average of similarity scores among the three profiles is expressed as

$$\overline{E}_{f,s,k} = \frac{\sum_{a=1}^{l} E_{f,s,k,fn}}{l}, \qquad (5)$$

where *f* is the phrase ID, *fn* is the number of elements in the phrase, and *l* is the number of profiles (in this case, *l*=3). We calculate $\overline{E}_{f,s,k}$ between each of the three standards and each phrase. Then, the standard that provides the lowest $\overline{E}_{f,s,k}$ is the category of the current phrase on the labelling. For example, if $\overline{E}_{f,s,k}$ between the melody and a phrase is the lowest, the phrase is labelled as melody.

C: Production of backing parts on arranged sheet music

The method of producing the backing on arranged sheet music involves first extracting note rows from the phrases labelled as backing, where the phrases have several chords. The note rows are assumed to be important on the arrangement, where a number of musicians for the backing are necessary. For example, if three out of five musicians are assigned to play the backing, the method extracts three note rows in the order of lowest pitch, highest pitch, and second lowest pitch. The reason this order is used is based on musical theory that appreciates outer voices as musically important. The generated note rows are composed into sheet music as well as the melody and base, so that sheet music for the musicians specified by the user who wants to arrange is finally generated, as shown at the bottom centre of Figure 1.

D: Consideration of naturalness and easiness

To improve the naturalness of the generated sheet music, we conducted a phrase exchange among musicians. To maintain ease of playing, the method allocates rests on the sheet music.

The sheet music for musicians in an ensemble exchanges heading phrases between musicians, and the exchange process is for creating musical dialogue. Two phrases, whose beginning times are synchronized and whose ranges of notes are within the range of the two musicians, are candidate of exchange. The proposed method is designed to ask the user who wants to arrange to determine whether the two have been exchanged.

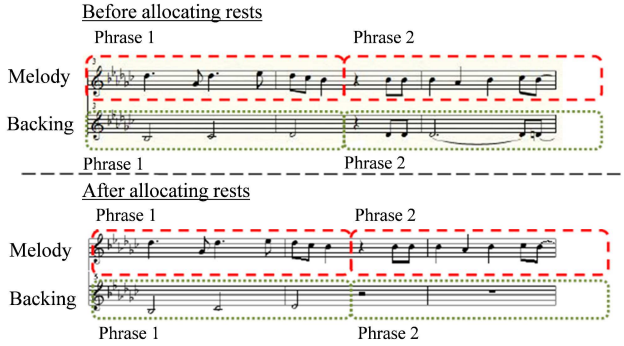
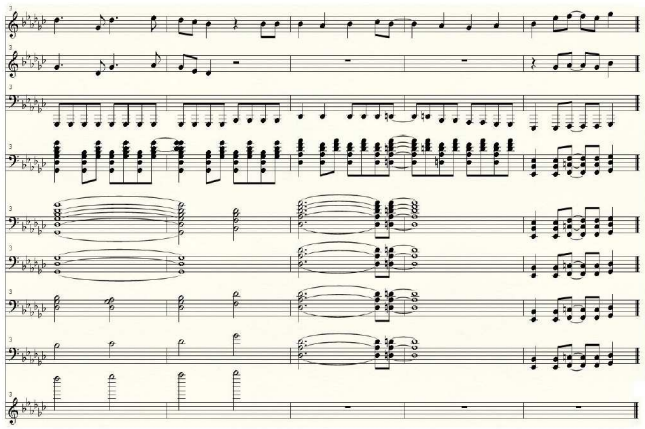
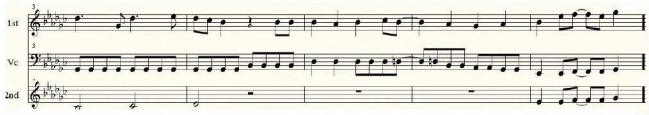


Figure 4: Example of allocating rests at phrases on backing part.



(a) Original sheet music



(b) Arranged sheet music (for 3 musicians)

Figure 5: Example of automatic arrangement generated using proposed method.

Also, to improve the ease for musicians, we allocated rests in several phrases. The tone rows of the parts (labelled as "Musician A" etc… in Figure 1) besides melody are chopped into phrases based on Miura's method. When the ending time of a phrase on the backing part synchronizes to that of the phrase on the melody part, the following phrase on the backing part is replaced with rests. Figure 4 shows an example of allocating rests in a phrase on the backing part. Phrase 2 of the backing part is replaced with rests.

Finally, we obtain the arranged sheet music by using the proposed method. Figure 5 shows an example of the automatically arranged sheet music generated using the proposed method.

4. INVESTIGATION OF DATABASE

A: Method of investigation

To evaluate the performance of the proposed method for extracting backing among phrases, we conducted an evaluation experiment to infer whether the proposed method can correctly determine the backing track. Eight excerpts were randomly chosen from thousands of MIDI excerpts, and for all eight MIDI excerpts, all tracks within a bar were extracted for rating by ten musicians who attend our university. Their ages ranged from 18 to 22, and all had more than two years musical expertise and are members of our university orchestra. They were asked to extract phrases of backing from the given phrases. We then selected the phrases of which more than seven of the musicians agreed were correct backing phrases. Two conditions of automatic labelling were used: "Condition Backing.", which is the proposed method, and "Condition ALL", which uses the Eingenphrase generated by all the tracks in the 4147 MIDI excerpts. By comparing the results of Conditions Backing. and ALL, it was possible to confirm the effectiveness of using the Eigenpharase of backing out of the Eigenphrase of all tracks.

B: Result of investigation

Figure 6 shows the results of the percentages of correct answers for both conditions. The correct ratio is the ratio of the number of answers that agree to the correct tracks to the number of all answers. The correct ratio of labelling based on the Eigenphrase of backing was 77% for Condition Backing, whereas those based Condition ALL was 56%. From the results of *t*-test of two samples, a significant difference (p<.05) between Cond. Backing and Cond. ALL was confirmed. Therefore, phrase labelling with our method was confirmed to be effective compared to the method that uses the Eigenphrase of all tracks.

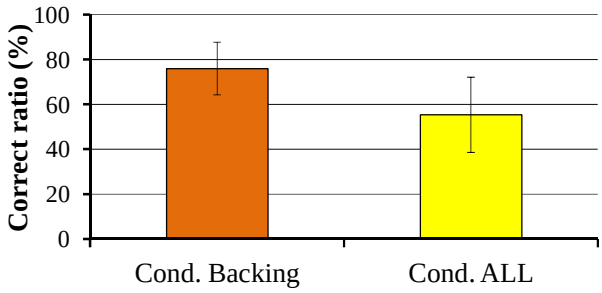


Figure 6: Correct ratio of classifying tracks by using two Eigenphrases. The ticks in figure represent 95% CI.

5. SUBJECTIVE EVALUATION OF ARRANGED SHEET MUSIC

A: Method of investigation

We conducted an experiment to investigated the effectiveness of automatically arranged sheet music generated using the proposed method in terms of musician impression. Three musicians (arrangers) and four other musicians (players) participated in our experiment. The three arrangers were asked to arrange four excerpts comprising 10 to 20 bars and the arranged sheet music was used. The four players were asked to play the sheet music produced with the following methods.

- Method P: Arrangement from the proposal method
- Method R: Arrangement by randomly selecting tracks
- Method H: Arrangement by a musician, who has experience in playing the violin
- Method S1: Arrangement by an expert percussionist
- Method S2: Arrangement by an expert in composition and arrangement

The four players were asked to play the arranged sheet music in random order and then rate their impression in terms of naturalness and easiness.

B: Results

Figure 7 shows the average and 95% confidence interval (CI) of naturalness scores, and Figure 8 shows the average and 95% CI of easiness scores. Figure 7 shows that Method P was rated higher than Method R for naturalness (p < .05 using the multiple test of Steel-Dwass), whereas there was no significant difference (p<.05) found between Method P and Methods S1, S2, and H. In addition, Method P was rated higher than method R in easiness.

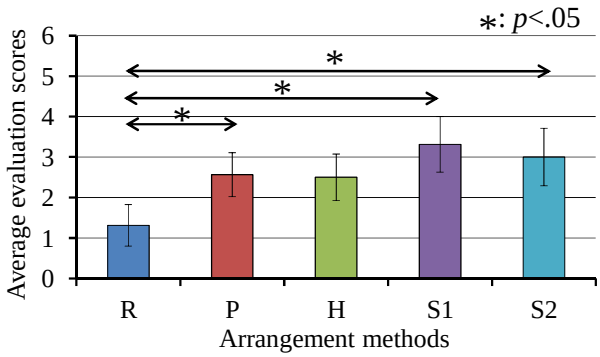


Figure 7: Average naturalness scores of arrangements with five methods.

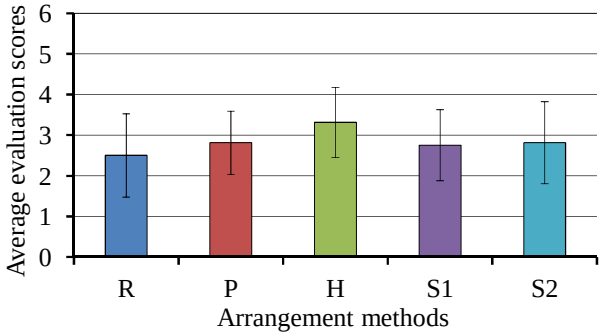


Figure 8: Average easiness scores of arrangements with five methods.

6. DISCUSSION

From the results of Methods P and R on naturalness, Method P generated more natural sheet music than Method R. No significant difference was found between the two methods regarding easiness. Interestingly Method H was rated the highest for easiness, implying that the arrangement by music experts was natural, but not always easiest to play. A possible reason for this is that the arrangement by experts may require a certain technical and complicated performance. There was no significant difference between Methods P and R, but Method P was higher than method R. The proposed method has the potential to generate sheet music that is easier to play than that through random selection.

7. CONCLUSIONS

We proposed an automatic arrangement method that takes into account naturalness and easiness. For naturalness, we used the Eigenphrase to correctly measure the similarity among phrases. To evaluate the effectiveness of the proposed method, we measured the percentage of correct answers to extract backing from phrases and found that the condition using the Eigenphrase of backing among all phrases was confirmed to extract backing from the phrases. For subjective evaluation of automatically arranged sheet music, we asked musicians to evaluate randomly selected arrangement excerpts, those by experts, and those with the proposed method in terms of naturalness and easiness. The proposed method maintained naturalness better than the method of random selection. Moreover, easiness of playing sheet music arranged using the proposed method and that by experts were almost the same; no significant difference was confirmed. For future work, we will discuss other aspects that affect ease of playing. The necessity of easiness with regards to arranged sheet music is still under discussion. If functions that accept users input for skill and to reflect that skill in arranged sheet music, easiness should be better incorporated into our method. Moreover, methods that do not involve observing the standard phrase of melody, base, and backing are preferable for more appropriate automatic arrangement.

8. ACKNOWLEDGEMENTS

We express our sincere thanks to the musicians of the Ryukoku University Orchestra for their help. This study is partly supported by the Grant-in-Aid for Scientific Research (25580050).

9. REFERENCES

Abe, Y., Murakami, Y., & Miura, M. (2012). Automatic arrangement for the bass guitar part in popular music using principle component analysis. Acoustical Science and Technology, *33, 4, 229-238.*

Miura, M., Inoue, Y., Murakami, Y., & Abe, Y. The automatic arrangement system and the method of automatic arrangement. Japan patent JP 2010-017807, (January 29, 2010).

Shakoh, S., Mizuno, R. & Kitamura, T. (2010). Automatic Arrangement System for Ensemble by Assign Melody and Genarate Accompaniment and Sub-melody. Information Processing Society of Japan, *2010-MUS-85, 1-6.*

Maekawa, H., Emura, N., Miura, M. & Yanagida, M. (2006). On machine arrangement for smaller wind-orchestras based on scores for standard wind-orchestras. Proc. of ICMPC9, *278-283.*

Yamaha Corp. (1995). Specifications of XF format (V2.03), *Retrieved November 22, 2000, from http://www.yamahasynth.com/jp/library/lectures/pdf/xfspc.pdf.*

Murakami,Y. (2009). Automatic arrangement for the drums-part based on case database for popular music, Master thesis, Ryukoku University, Shiga, Japan.

Nagasima, Y., Hashimoto, S., Hiraga,J., & Hirata, K. (1999). The world of computer and music *(pp. 224-228), publisher of Kyouritu.*

[1-10] ERROR CULTURE IN INSTRUMENT LESSONS IN SOUTH KOREAN MUSIC SCHOOLS

Sun Hee Lee[1]
[1]*University of Music and Performing Arts Graz, Austria*
Poster

Errors occur naturally in the learning process; according to pedagogy theory, errors can be regarded as a learning potential. During an instrument lesson, there are likely to be many such errors. Usually, these are immediately corrected by the teachers in order to prevent the fault from being engrained in the student's memory. Such quick correction by teachers though can have a negative influence on a student, reflecting an inefficient awareness of the important role that errors play in learning. How teachers deal with errors is one of the fundamental components of teaching expertise. Furthermore, the error culture is related to its philosophic-historical background (its context in the history of ideas) so its nature may be different in South Korea. The purpose of this dissertation, first of all, is to explain about error culture in the pedagogical field related to instruments and to investigate the background of error culture. The aim of this study is to sharpen the awareness of dealing with errors and to present this issue more transparently, so that errors can be constructively integrated into the learning processes. Second, this research as a comparative study intents to investigate the particular factors influencing the error culture and instrument lessons in South Korea. This study focuses on the following research questions, linked to the aims of this study: (1) What does the error culture in instrument lessons in South Korea look like? The following additional questions will also be investigated: 1) How do instrument teachers in music schools deal with students' errors? 2) In what ways is learning from errors possible? 3) What approaches are there in order to productively deal with errors during instrument lessons? (2) How does the intellectual context (context in the history of ideas) affect the error culture? The method for this study is a qualitative method, using interviews. The survey subjects will be Korean piano teachers in Seoul.

[1-11] THE EFFECT OF MOTOR INVOLVEMENT AND MELODY TRUNCATION ON INVOLUNTARY MUSICAL IMAGERY

Stephanie Mccullough[1]
[1]*University of Arkansas, USA*
Poster

1. Background: The term "earworm," also known as Involuntary Musical Imagery or INMI, refers to the phenomenon of an uncontrollably repeating melody in one's head. Though ubiquitous, it is comparatively under-researched in music cognition. Most existing studies have catalogued descriptive factors. This research has successfully identified the defining characteristics of earworms, but almost none of it explores underlying mechanisms. This study addresses that gap by investigating a specific theory about INMI induction and examining factors that influence earworms' "stickiness."

2. Aims: I plan to investigate the hypothesis that overt motor involvement (in the form of humming, singing, tapping or moving) and imagined motor involvement (in the form of imagining a continuation to an interrupted melody) will induce INMI more frequently than passive music listening.

3. Method: Four groups of participants complete a monotonous visual activity while listening to music. The goal of this activity is to achieve a low attention state for participants; Williamson et al. (2011) found this primed subjects for earworm induction.The control group hears a song that ends at a normal phrase ending; the second group hears a song that stops abruptly in the middle of a phrase. The third group engages with the music by humming, whistling, or singing along. The fourth group engages kinetically by dancing, tapping, or nodding. Groups complete the same mundane visual task in each condition. After the music and task simultaneously end, participants complete a basic demographic questionnaire while sitting in silence. Finally, they are asked to report on any earworms that occurred during the session.

4. Results: I hypothesize that engaging with the music motorically, whether overtly or via imagination, will recruit the cortico-basal ganglia network that makes earworm induction more likely.

5. Conclusions: This experiment sheds light on a proposed mechanism for earworm induction, and it is relevant both to basic science about the structure of the auditory system and to applied concerns about controlling earworms.

[1-12] MODULATING ATTENTION IN MULTI-PART MUSIC: IS SIMPLE CONSONANCE MORE HELPFUL THAN COMPOUND DISSONANCE?

Rhimmon Simchy-Gross[1], *Elizabeth Margulis*[2]
[1]*Department of Psychological Science, University of Arkansas, USA,* [2]*Department of Music, University of Arkansas, USA*
Poster

1. Background: Collaborative musicians necessarily use prioritized integrative attention (PIA), dividing attention between their own part and the aggregate texture in real time, to synchronize their actions and create a unified musical entity (Keller, 2001). In a pioneering study, Keller and Burnham (2005) measure rhythm recognition accuracy to indicate attentional focus and provide evidence that metric compatibility facilitates PIA. Yet, despite the role that PIA holds for successful coordination in joint music-making activities (e.g., ensembles, choirs, etc.), empirical investigations of the factors that modulate PIA remain rare.

2. Aims: This study aims to show how intervallic stability and distance modulate PIA. The dynamic attending theory (Jones & Boltz, 1989) and the theory of auditory stream segregation (Bregman, 1990) shape our predictions, where harmonically congruent intervals establish 'harmonic markers' and large-sized intervals facilitate voice separation, respectively.

3. Method: Two alternating tones (trumpet and piano timbres) were characterized by Consonant (unison) or Dissonant (minor 2nd) and Simple (no-octave) or Compound (two-octave) intervals. Accordingly, four conditions (Simple-Consonant; Compound-Consonant; Simple-Dissonant; Compound-Dissonant) were randomly varied within subjects. Subjects (listeners with formal musical training) memorized the rhythm of the "part" (trumpet) while simultaneously memorizing the rhythm of the "whole" (trumpet and piano together). Then, after hearing a similar or identical rhythm (for either "part" or "whole"), subjects rated their level of confidence that the corresponding rhythms were in fact the same or different using a 6-point Likert-like scale.

4. Results: We predict a main effect of stability that is, a processing advantage for both rhythms ("part" and "whole") in the Consonant (versus Dissonant) conditions. We furthermore predict a distance-by-rhythm interaction that is, a processing advantage for "part" in the Large (versus Small) and for "whole" in the Small (versus Large) conditions.

5. Conclusions: We suspect that intervallic stability facilitates PIA whereas, intervallic distance (i.e., large and small) facilitates the individual properties that constitute PIA (i.e., selective and integrative attention, respectively). Taken together, these data can inform music-listening strategies, pedagogical tools, therapeutic practices, and performing techniques by providing evidence for the effects of two unique intervallic properties (stability and distance) on some of the attentional mechanisms involved in music perception and performance.

[1-13] THE EFFECT OF MUSIC TEMPO AND BACKGROUND NOISE ON ADULTS' READING COMPREHENSION AND READING SPEED

Philip Fine[1], *Lucy Kirby*[1]
[1]*University of Buckingham, United Kingdom*
Poster

ABSTRACT

The current study examined the effect of background music and noise, and of musical tempo on adults' reading comprehension ability and reading speed. Forty-eight adults carried out reading comprehension tasks under four conditions: fast-tempo music, slow-tempo music, background noise, and silence. The results showed an effect of condition on reading speed but not on reading comprehension. The passages were found to differ in terms of their difficulty level, irrespective of condition. How often participants listen to music did not affect reading comprehension or distraction ratings, though higher distraction ratings correlated with poorer reading comprehension. Participants found silence less distracting than the other three conditions. The results of the present study are of interest to, inter alia, students who choose to study with music or background noise and also for educators when selecting background music or noise for learning situations.

1. INTRODUCTION

People often carry out tasks in the presence of background music or radio/TV. Background music is common in many settings in today's society, including those relating to the workplace, health, education, marketing and retail (Doyle & Furnham, 2012). Many students study whilst listening to music, others preferring to study in silence (Adriano, 2010). Music helps some students focus, whereas to others, it is a distraction: this may depend on extraversion level (Geen, 1984). A survey of 1700 Dutch students found that 80% regularly did their homework with the radio on and 50% with the TV on in the background (Beentjes, Koolstra, & van der Voort, 1996). Interestingly, this was the case even though the students generally felt that their assignments were of poorer quality if they worked with background music. Various factors mediate the effect of background music and noise on task performance, including the type of music or noise, individual differences of the listener (e.g. personality), the task being carried out, and the level of attention paid to the music or noise.

Cognitive psychologists study how music affects performance on various cognitive tasks, such as reading and writing, which are both highly relevant to studying. Undergraduates' ability to type fluently was significantly disrupted by listening to background music (Ransdell & Gilroy, 2001), perhaps due to the high demand music places on working memory, even when unattended. The presence and type of background music significantly affects reading rate and emotional evaluation of news read on a pocket computer in a busy environment (Kallinen, 2002). Furnham and Strbac (2002) found that reading comprehension, memory and arithmetic tasks were all impaired in the presence of background music and noise compared to silence. The effect of irrelevant noise on cognitive performance has also been investigated. Background noise can act as a distracter and impair reading comprehension compared to silence (Furnham, Gunter, & Peterson, 1994; Furnham & Strbac, 2002), though it did not differ in its effect from background music.

Individual differences moderate the use (Geen, 1984) and effects of background music on cognitive performance. For example, creative individuals tend to listen more to music when studying and are less distracted by it compared to non-creative individuals (Doyle & Furnham, 2012). Introverts are more affected by background noise and music than extraverts (Dobbs, Furnham, & McClelland, 2011; Furnham and Bradley, 1997) when carrying out cognitive tasks such as reading for comprehension. Background music does therefore seem to affect cognitive processing, but these effects differ between individuals.

The music that students listen to when studying varies greatly. Whilst some listen to instrumental or classical music, others listen to pop or heavy metal. Complex vocal music impairs task performance more than less complex instrumental music (Furnham & Bradley 1997). Aspects of the music itself, such as tempo and intensity, can also moderate its effects when in the background. For instance, fast tempo, high arousal music disrupts memory and reading comprehension tasks more than slow tempo, low arousal music (Cassidy & MacDonald, 2007; Thompson, Schellenberg, & Letnic, 2012). The lower the tempo, the more positive its effect on both reading comprehension and mental arithmetic (Furnham & Strbac, 2002).

However, some studies have suggested that fast tempo can enhance cognitive processing. For instance, Angel, Polzella, and Elvers (2010) showed that fast tempo background music increased both the speed of spatial processing and the accuracy of linguistic processing. Kallinen (2002) found that reading efficiency and rate was significantly worse in the presence of the slow music compared to fast music. Conversely, a meta-analysis by Kämpfe, Sedlmeier, and Renkewitz (2010) found no overall effect of tempo on cognitive performance, but did show that fast tempo can increase the speed at which behavior, such as eating, is carried out, even without conscious awareness. Kämpfe et al. also found that the presence of background music disturbs the reading process and has a negative effect on memory. On the other hand, their meta-analysis showed that music can have a positive influence on emotional reactions and improve people's sporting performance. Thus the effect of music and music tempo on cognitive functioning is complex and multidimensional.

Background music can have a positive effect on certain tasks, particularly repetitive ones (Fox & Embrey, 1972), suggesting that the effect of background music or noise may be task dependent. However, Etaugh and Ptasnik (1982) demonstrated that music had a positive effect on literacy tasks such as reading comprehension, but only for students who often work with music: those who normally worked in silence were impaired by background music, suggesting the importance of individual differences. However, Furnham and Bradley (1997) concluded that there is little evidence that the presence of background music or noise is beneficial to performance on complex cognitive tasks.

The current literature thus shows mixed findings, with background music shown to be either beneficial to, detrimental to, or to have no effect on task performance. These inconsistencies may relate to several methodological factors, such as task type, music genre, and participant groups. The literature concerning how music tempo and background noise affect reading comprehension is also inconsistent. The current study aims to investigate the impact of both background music (and its tempo) and background noise on reading performance in the same group of participants. As many students study with background music or noise (whether through choice or not), it is pertinent to ask what effect this has on study-related cognitive abilities such as reading comprehension.

Based on previous research four predictions were made: a) reading comprehension scores would be highest in the silence condition due to less distraction and interference with memory processes (Furnham & Strbac, 2002); b) reading speed would be faster in the fast tempo condition compared to the slow tempo condition (Kämpfe et al., 2010); c) those who more often work with music would find music less distracting than participants who do so rarely (Furnham & Strbac, 2002); and d) people who more often listen to music while working would score better on the reading comprehension in both music conditions compared to participants who rarely listen to music (Etaugh & Ptasnik, 1982).

2. METHOD

2.1. Participants and design

Forty-eight participants (18M, 30F) took part. Their ages ranged from 19 to 71 years (M = 24.8, SD = 10.3). All had English as their first language or were bilingual, and none had language difficulties such as dyslexia. A within-subjects design was used: each participant was tested under each of the four conditions. The independent variable was condition (fast music, slow music, background noise, silence). The main dependent variables were reading comprehension score, reading speed, calculated in wpm (words per minute), and time spent answering the questions. Distraction ratings for each condition were also recorded.

2.2. Materials

Four short passages from the Adult Reading Test (Brooks, Everatt, & Fidler, 2004) were used. Each passage was followed by ten questions based on that passage. The passages covered the topics of Health, News, Business and Film respectively. The four passages used increased in difficulty (from A to D), so their order was partially counterbalanced across participants. Four 'easy-listening' songs chosen from the Hôtel Costes volumes were used. The slow tempo music (84 bpm beats per minute) comprised 'Mirage (feat. Michael Robinson)' by Shazz and 'Dead End Street (Trentmøller Remix)' by Morten. The fast tempo music (130 bpm) comprised 'The Game (feat. Ashley Slater)' by Dublex Inc. and 'To Ulrike M (Zero 7 Mix)' by Doris Days. Complex vocal music was chosen based on previous research findings (Furnham & Bradley, 1997). Suitable background noise was found on YouTube (http://www.youtube.com/watch?v=Cny-s1oKZ7M): speech with no intelligible words. The music was played through either Bose QuietComfort noise cancelling QC15 headphones or Bose OE audio headphones via a Macbook Pro laptop at volume level 7, so volume level was kept consistent for all participants.

2.3. Procedure

Participants were tested individually. After giving informed consent, they provided demographic data, including age, gender, and how often they worked with background music and noise, rated on a 7-point Likert scale (Furnham & Strbac, 2002). They then completed the reading comprehension tasks, one in each condition. After reading each passage straight through once at their own speed, they answered the comprehension questions without being able to refer back to the passage. Reading time (to calculate reading speed) and time to answer the questions were both recorded for each passage. Participants wore headphones for the whole task (both reading the passage and answering questions) as the music or noise was played throughout. They also wore headphones during the silent condition for consistency and to limit any background noise during testing. After completing each reading comprehension task, participants rated on a 7-point Likert scale how distracting they found the condition. At the end of the study, participants were asked whether they recognized any of the music used (none did). The order of the conditions was fully counterbalanced to control for practice and boredom effects.

3. RESULTS

Table 1 shows reading comprehension scores, reading speed, and time taken to answer the questions, for the four conditions: fast tempo music, slow tempo music, noise, and silence.

	Fast	Slow	Noise	Silence
Comprehension	5.5 (2.6)	5.7 (2.1)	6.1 (2.3)	6.2 (2.1)
Reading speed	178 (65)	168 (62)	185 (75)	187 (77)
Answering time	161 (70)	163 (71)	157 (71)	174 (67)

Table 1: Mean (SD) reading comprehension scores (out of 10), reading speed (wpm) and time taken (seconds) for each condition.

Background music appears to reduce reading comprehension, but repeated measures ANOVAs demonstrated that condition did not significantly affect reading comprehension, reading speed or answering time (p>.01). However, an analysis of covariance controlling for how often participants listen to music whilst working revealed that condition significantly affected reading speed ($F_{3,138}$=5.87, p=.001), slow tempo reading speed being slower than fast tempo reading speed ($F_{1,46}$=8.64, p=.005).

Table 2 shows reading comprehension scores, reading speed, and time taken to answer the questions, for the four passages used.

	A	B	C	D
Comprehension	7.1 (1.7)	5.4 (2.2)	6.3 (2.2)	4.7 (2.2)
Reading speed	215 (69)	133 (50)	186 (69)	182 (64)
Answering time	152 (53)	155 (68)	196 (83)	146 (63)

Table 2: Mean (SD) reading comprehension scores (out of 10), reading speed (wpm) and time taken (seconds) for each condition.

Repeated measures ANOVAs demonstrated that the passages differed significantly in comprehension score ($F_{3,141}$=20.22, p<.001), reading speed ($F_{2.27,106.51}$=42.14, p<.001, Greenhouse- Geisser correction used), and question answering time ($F_{3,141}$=21.55, p<.001). Reading

comprehension was significantly worse for passages D and B than A (p<.001). Reading speed was significantly slower for passage B than all other passages (p<.001), and significantly faster for passage A than all other passages (p<.01). Time to answer the questions was significantly longer for passage C than all other passages (p<.001).

Average distraction scores (on a 1-7 scale, 7 most distracting) for the four conditions are shown in Table 3. Participants rated silence as significantly less distracting ($F_{3,138}$=30.20, p<.001) than all other conditions. An analysis of covariance revealed that the effect of condition on distraction was still significant ($F_{3,138}$=13.43, p<.001) when we controlled for how often participants listen to music whilst working (rated on a 1-7 scale).

	Fast	Slow	Noise	Silence
Distraction	5.1 (0.2)	4.4 (0.2)	4.6 (0.2)	2.3 (0.2)

Table 3: Mean (SD) distraction ratings for each condition.

The more often participants worked with music (rated on a 1-7 scale), the more distracting they found the silence condition (r_{48}=0.29, p<0.05), but there was no relationship between preference for working with music and distraction ratings for the other three conditions. How often participants listened to music whilst working did not correlate with either overall reading comprehension or distraction ratings. However, there was a significant correlation between reading comprehension performance and mean distraction ratings for all conditions except silence (r_{48}>0.33, p<=.02): the more distracting the background sound, the worse the participants' comprehension performance.

We calculated total task time by adding reading time and answering time. Comprehension score correlated significantly with both answering time (r_{48}>.29, p<.05) and total task time (r_{48}>.30, p<.04) for all conditions except silence: the longer the participant spent answering the questions or overall on the task, the better they did. However, reading speed did not correlate with comprehension scores in any condition. Finally, there were no gender differences on any variable.

4. DISCUSSION

The current findings suggest that reading comprehension is unaffected by the presence of background noise or music, or its tempo. This is somewhat surprising, as it does not support existing research showing impaired reading comprehension with background sound compared to silence (Furnham & Strbac, 2002; Kallinen, 2002; Kämpfe et al., 2010). The results partially support Kämpfe et al.'s findings that music tempo does not affect cognitive performance but does appear to affect reading speed when we control for how often people work to music. However, reading speed was slower with background music than either sound or silence. Kallinen (2002) suggests that using headphones may moderate the effect of background music on reading comprehension, in that they isolate the listener from other audible distractions in the environment. In this study, noise-cancelling headphones were used, further isolating the participants. In daily life, those studying with music may not use headphones, and thus other sounds in the environment will add to the distraction. Other explanations might relate to the specific task or to participants' personality or motivation.

The four passages (Brooks et al., 2004) were graded and known to differ in difficulty. This was clearly demonstrated in significant differences in reading comprehension, reading speed and time taken to answer the questions for the four passages. Each passage concerned a different topic, and it is possible that some may have interested and motivated certain participants more, affecting both their reading speed and reading comprehension. The differences in topic and difficulty are clearly limitations to the present study, but counterbalancing the order of passages and which passage was combined with which condition should have minimised such effects. In future, interest ratings in each passage should also be recorded to investigate their influence.

Participants were asked to rate how distracting they found each condition. Silence was deemed significantly less distracting than the other conditions, even when preference for working with background music or noise was controlled for. However, those who tended to work more with background music or noise found silence more distracting, suggesting that they find the lack of environmental noise off-putting (Adriano, 2010; Furnham & Strbac, 2002), indicating the importance of what one is used to. For each condition other than silence, participants' reading comprehension scores correlated negatively with how distracting they found the condition: even though distraction level did not differ overall between non-silence conditions, participants tended to perform worse when they were more distracted.

There was no association between how often participants work while listening to music and their reading comprehension scores or distraction ratings, suggesting that even though some choose to work with background sound, this does not affect their reading comprehension ability. The presence of background music or noise was rated as more distracting than silence overall, even though it did not seem to affect reading comprehension scores or reading speed. Again, this may relate to individual differences between the participants, such as level of introversion (Dobbs, Furnham, & McClelland, 2011; Furnham & Bradley, 1997; Furnham & Strbac, 2002), although this was not measured in the present study. Such measurements would be a good idea in future studies. Finally, reading comprehension performance correlated with how long it took the participants to answer the questions but not reading speed. Perhaps this was merely because it takes a roughly constant time to answer each question, and attempting more questions tended to take longer and result in a higher score.

The study had a number of limitations which any replications could address. The passages varied in difficulty, significantly affecting comprehension score, reading speed and total task time. Counterbalancing should have addressed this and hence any effects of condition on reading comprehension and speed should not have been masked, but it would be sensible to replicate the study with four text passages of equivalent difficulty. Participants should also rate their interest in each topic so that motivational influences could be accounted for. The reading comprehension task used in the study was unlike students' typical work assignments, and a range of more realistic assignment tasks could be used. Also, participants were tested on one passage straight after another, so fatigue or boredom could have influenced the findings. Again, counterbalancing the order of conditions should have addressed this at least to an extent. Finally, no information on individual differences in terms of personality (such as introversion level) or music preference (such as genre) was recorded, and this may have an effect on the complex interplay between the background music or noise and cognitive task performance.

These findings illustrate the multi-dimensional nature of the effects of background music and noise on reading. In particular, both background music and noise were rated as more distracting than silence, and people who were more distracted by background sound had poorer reading comprehension. This suggests that students should, where possible, make realistic choices as to whether to study with background music or radio/TV, and if they find it distracting, to turn it off. Future studies can investigate how individual differences and music-related aspects such as

genre or familiarity might mediate the effects of background sound on both cognitive ability and distractibility, specifically in tasks more representative of school or university study.

5. REFERENCES

Adriano, J. (2010). Teenagers' reasons for listening to music and the students' perception of the effects of listening when completing school assignments. Unpublished thesis, Johnson and Wales University.

Angel, L. A., Polzella, D. J., & Elvers, G. C. (2010). Background music and cognitive performance. Perceptual Motor Skills, 110, 3, 1059-1064.

Beentjes, J. W. J., Koolstra, C. M., & van der Voort, T. H. A. (1996). Combining background media with doing homework: Incidence of background media use and perceived effects. Communication Education, 45, 59–72.

Brooks, P., Everatt, J., & Fidler, R. (2004). Adult Reading Test. Pearson Clinical.

Cassidy, G., & MacDonald, A. R. (2007). The effect of background music and background noise on the task performance of introverts and extraverts. Psychology of Music, 35, 517-537.

Dobbs, S., Furnham, A., & McClelland, A. (2011). The effect of background music and noise on the cognitive test performance of introverts and extraverts. Applied Cognitive Psychology, 25(2). 307-313.

Doyle, M., & Furnham, A. (2012). The distracting effects of music on the cognitive test performance of creative and non-creative individuals. Thinking Skills and Creativity, 7(1), 1-7.

Etaugh, C., & Ptasnik, P. (1982). Effects of Studying to Music and Post-study Relaxation on Reading Comprehension. Perceptual and Motor Skills, 55, 141-142.

Fox, J. G., & Embrey, E. D. (1972) Music—An aid to productivity. Applied Ergonomics, 3(4), 202-205.

Furnham, A., & Bradley, A. (1997). Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts. Applied Cognitive Psychology, 11, 445-455.

Furnham, A., Gunter, B., & Peterson, E. (1994). Television distraction and the performance of introverts and extraverts. Applied Cognitive Psychology, 8, 705-711.

Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: the differential distraction of background music and noise on the cognitive test performance of introverts and extraverts. Ergonomics, 45(3), 203-217.

Geen, R. G. (1984). Preferred stimulation levels in introverts and extroverts: Effects on arousal and performance. Journal of Personality and Social Psychology, 46(6), 1303.

Kallinen, K. (2002). Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. Computers in Human Behaviour, 18, 537-551.

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. Psychology of Music, 39(4), 424-448.

Ransdell, S. E., & Gilroy, L. (2001). The effects of background music on word-processed writing. Computers in Human Behavior, 17, 141-148.

Thompson, W. F., Schellenberg, E., & Letnic, A. K. (2012). Fast and Loud Background Music Disrupts Reading Comprehension. Psychology of Music, 40(6), 700-708.

[1-14] DEVELOPMENT AND VALIDATION OF THE INVOLUNTARY MUSICAL IMAGERY SCALE (IMIS)

Georgina Floridou[1], Lauren Stewart[1], Victoria Williamson[2,3], Daniel Müllensiefen[1]
[1]Department of Psychology, Goldsmiths, University of London, United Kingdom, [2]Hochschule Luzern – Musik, Lucerne University of Applied Sciences and Arts, Switzerland, [3]Department of Music, University of Sheffield, United Kingdom

Poster

1. Background: We report on the development of a novel self-report instrument, the Involuntary Musical Imagery Scale (IMIS), which measures different aspects of an involuntary musical imagery experience (more commonly termed an 'earworm'). The new scale compliments existing measures and comes with extensive psychometric testing and validation against related aspects of cognitive behavior and psychological traits such as musical behaviors, auditory imagery abilities, low attention traits and cognitive intrusions.

2. Aims: The first aim of the study was to develop a scale to measure individual differences relating to different aspects of the INMI experience. The second aim was to look into the relationship of the scale factors with other related phenomena and existing scales.

3. Method: Exploratory factor analysis was conducted on the data from 362 participants who completed a preliminary online version of IMIS (68 items). Using exploratory factor analysis and psychometric assessment methods, the number of items was considerably reduced (15 items) and the factor-structure was revealed. For the subsequent confirmatory analysis, 612 participants have so far completed the IMIS. In addition they have completed a number of additional scales regarding musical behaviors, auditory imagery abilities, low attention traits (such as tendency to daydream) and cognitive intrusions. Finally, a pen and paper version of IMIS was completed by 147 psychology undergraduate students along with the Bucknell Auditory Imagery Scale (BAIS).

4. Results (to date): Exploratory factor analysis revealed 4 factors, namely Negative Valence, Movement, Personal Reflections and Help. Confirmatory factor analysis is in progress. The pen and paper version revealed interesting positive correlations between INMI frequency and the Vividness scale (BAIS) as well as all the IMIS factors.

5. Conclusion: The newly developed IMIS appears to be a valid and reliable tool for the evaluation of INMI experiences and behaviours. Further analysis will provide useful insights regarding the relationship of the IMIS factors, frequency and duration with related phenomena, as well as convergent and divergent validity in comparison with other instruments for the first time in the exploration of INMI.

[1-15] REPORTED PLEASANTNESS OF INVOLUNTARY MUSICAL IMAGERY CORRELATES WITH GRAY MATTER DENSITY IN EMOTION-RELATED AREAS

Nicolas Farrugia[1], Kelly Jakubowski[1], Robert Carlyon[3], Rhodri Cusack[2], Lauren Stewart[1]
[1]Goldsmiths, University of London, London, United Kingdom, [2]Brain and Mind Institute,western University, London, Ontario, Canada, [3]Mrc Cognition & Brain Sciences Unit, Cambridge, United Kingdom

Poster

1. Background: Involuntary Musical Imagery (INMI), or earworms, is a short section of imagined music that comes into the mind and repeats involuntarily. Recent research has shown that the subjective experience of INMI is variably reported as pleasant, neutral or irritating, suggesting a role of individual differences in the emotional response to INMI.

2. Aims: In the present study, we test the hypothesis that the self-reported valence of INMI experience is associated with structural variation in cortical areas linked to emotion processing.

3. Methods: 43 healthy participants (aged from 20 to 75 years old) used a web interface to complete the IMIS (Involuntary Musical Imagery Scale), a questionnaire developed to probe subjects about their subjective experience of INMI in daily life, which includes a subscale focusing on the perceived valence of their experiences (e.g. how much individuals were irritated, worried or would like to suppress INMI episodes). T1-weighted images were acquired using an MPRAGE sequence on a Siemens 3T Tim Trio. Whole-brain analysis of gray matter density was performed using Voxel Based Morphometry (VBM) and DARTEL. Local gray matter was analyzed using a linear model with negative valence as a covariate of interest, and age, gender and total gray matter as nuisance parameters. Reported p-values relate to effects at the cluster level, corrected for Family-Wise Error rate and non-stationary cluster extent.

4. Results: INMI negative valence scores were positively related to local gray matter volume in right temporopolar cortex ($p < 0.005$), right orbitofrontal cortex ($p = 0.1$), and left temporopolar cortex ($p = 0.1$). Post-hoc meta-analysis of previous fMRI studies using Neurosynth associated these cluster locations to features such as negative emotion, unpleasant experience, distress, disgust or reward.

5. Conclusions: These results suggest a role of deep cortical and subcortical emotional networks in the response to INMI, indicating that subjective experience of INMI may rely on neural mechanisms responsible for basic emotions. We will extend this approach to study the structural and functional networks involved in INMI.

[1-16] MAPPING MUSICIANS' MEMORY

Stefania Pileri[2], Cristina Di Bernardo[2], Alan D Baddeley[3], Graham J Hitch[3], Daniel Mullensiefen[2], Victoria J Williamson[1]
[1]University of Sheffield, United Kingdom, [2]Goldsmiths, University of London, United Kingdom, [3]University of York, United Kingdom
Poster

1. Background: Expertise is associated with improvements to memory that are specific to the practiced domain. Musical expertise is one skill where evidence has previously supported transfer to domain-general memory abilities, mostly with verbal stimuli. Although several studies of musicians have included more than one memory task, there has so far been no systematic comparison across multiple memory systems and therefore the precise pattern of potential transfer from musicianship across memory remains unknown.

2. Aims: In this pilot study we aimed to i) replicate findings regarding a relationship between musicianship and enhanced verbal long-term memory (LTM), ii) investigate the impact of musicianship on other forms of verbal memory (short-term memory (STM), working memory (WM)) and iii) compare memory abilities outside the auditory domain (visual, spatial).

3. Method: We used a battery of nine memory tests (11 test scores) to compare performance of 18 non-musicians and 18 musicians, matched on age, gender, education, employment and socio-economic status, and verbal intelligence. The tests were the Rey Auditory Verbal Learning Test, and the Wechsler paired-associate task (both verbal LTM), operation span (WM), forward, backward and combined Digit Span (STM), Corsi Spatial Span and Visual Pattern Span, and a published measure of musical STM.

4. Results: A series of t-tests found a significant effect of musicianship on 6 out of the 11 scores, including working memory and measures of verbal LTM. However, to account for the potential confounding influence from the other memory tests, we employed Rubin's causal analysis method. Here we found no effect of musicianship on memory performance with the exception of musical STM. Finally, to elucidate the dependence structure among the full set of memory tests we performed a causal network analysis. This preliminary analysis indicated a specific effect of musicianship on musical STM.

5. Conclusions: This pilot study provisionally confirms an association between musicianship and memory performance but at present the data are not sufficient to separate the influence on the specific aspects of memory tested. Further testing is planned and we discuss the importance of controlling for relevant intellectual differences between adult musician and non-musician groups.

[1-17] OCCURRENCE AND PROCESS OF THE PERFORMER'S ANXIETY FOR OPERA AND PIANO CONCERT BY UNIVERSITY STUDENTS

Yoko Ogawa[1]
[1]Okayama University, Japan
Poster

ABSTRACT

This study is to clarify how subjects, Japanese university students', feel and react to performance anxiety or stage fright. It is commonly said that some of us feel nervous when performing on stage, in front of an audience. Regardless of amateur or professional skill, every musician faces a variety of performance anxiety. We also know that performance anxiety is a phenomenon that is dependent on a variety of factors such as: audience, performers, instruments, situations, and context. Furthermore, the growth and process of each performance anxiety varies according to the situation, and can affect the same subject differently. The purpose of the present study was to explore the occurrence and process of specific performance anxiety by comparing 2 different formal situations: an Opera and an annual concert performance.
The research questions were: (1) When and how do university students feel anxiety throughout the Opera situation*?* (2) When and how do they feel anxiety throughout their concert performance*?* (3) Are there any differences or similarities between anxieties facing Opera and annual concert performers? (4) Are there plausible explanations for these differences, if any? In Experiment 1, 11 university students were asked to judge their feelings about their Opera performance on a STAI Y-1 (State-Trait Anxiety Inventory-Form) sheet: in 20 sentences using 4-point rating scales, and SD method: 5-point rating scales, whose ends were indicated by pairs of adjectives. In Experiment 2, the same participants were asked to judge their feelings throughout a regular concert performance, using the same procedure as Experiment 1. All student data were gathered and classified into 2 groups. Results of the Opera performance, using STAI Y-1, documented 2 patterns in the test profile. Some students who played

the cast roles showed that cheerfulness overcame their anxiety, while students in stagehand roles displayed stronger uneasiness up until performance day. Similar profile patterns were found between annual concert performers and Opera performers. Additionally, one factor labeling pleasant was identified in both situations.

1. BACKGROUND

We sometimes feel nervous when performing on stage in front of an audience. According to Wesner, Noyes & Davis (1990), 25 % of faculty students at the American University School of Music reported "marked distress", and 40% reported "moderate distress". A similar study suggested that 59% of professional orchestra members are affected by stage fright (Van Kemanade, et.al, 1995). Regardless of amateur or professional skill, every musician faces a variety of performance anxiety. We also know that performance anxiety is a phenomenon that is dependent on a variety of factors such as: audience, performers, instruments, situations, and context. It is said that the quality of performance is related to arousal with an inverted-U curve, Yerkes-Dodson Law. According to this law, very low levels of arousal are insufficiently motivating and give rise to lackluster performances, and excessive arousal also interferes with performance because concentration is disrupted, memory blocks occur, and there is a loss of steadiness in hands and voices (2002, Wilson & Roland). In other words, a certain degree of arousal actually helps the quality of performance, however, we often fail to grasp how degree of arousal is suitable for our performance. To find a way out of these difficulties, it is necessary to stack several experiments along some stages. Therefore, we would like to clarify how university students react to the performance anxiety under 2 different formal conditions as the first stage of this study.
Therefore, we would like to clarify how university students react to performance anxiety under 2 different formal conditions as the first stage of this study. The following research questions were asked:

(1) When and how do university students feel anxiety throughout the Opera situation*?*

(2) When and how do they feel anxiety throughout their annual concert performance*?*

(3) Are there any differences or similarities between anxieties facing Opera and annual concert performers?

(4) Are there plausible explanations for these differences, if any?

2. STUDY 1

2.1. Method

Subjects. The participants were 11 subjects: 4 males and 7 females. They were undergraduates and postgraduates at Okayama University, aged 20 to 24 years old (average age was 22.1yrs). Some participants were music education majors and others were music performance majors, and wanted to teach at high schools or junior high schools in the future. The length of their music training was over 10 years, and all of them were doing some kind of musical activity at the time of this experiment.

Procedure. Each subject was asked to judge their feelings about their Opera performance on a STAI Y-1 sheet: in 20 sentences using 4-point rating scales: 4 is strong anxiety, 1 is no anxiety. An opera performance titled "The marriage of Figaro" was presented on the 15th of Dec. 2012. Whole subjects prepared for this opera from Apr. 2012. Although the investigations were held 4 times: 3 months, 2 months, 1 months and one day prior to performance, we focused on the transition and the holding of the final rehearsal. Furthermore all subjects responded to the SD method in order to show their emotion at the final rehearsal. The adjective made 16 pairs in reference to the result of some previous studies and the preliminary experiment. In a rehearsal room, each subject was instructed to evaluate their feeling using 5-point rating scales, whose ends were indicated by pairs of adjectives. The entire test required approximately 15 minutes to complete.

2.2. Results and findings

All students' data from STAI Y-1 were gathered firstly and subsequently classified into 2 groups. As shown in the following 3 figures, some movements of subjects' anxiety tended to be reduced, while others did not. Specifically, we can detect a big difference between the 3rd and 4th investigation. Therefore, those without anxiety expressed at the 4th investigation were called the "anxiety free group", and the other group who demonstrated tension, were called the "anxiety group". According to one student's comment after investigations, who was categorized in the "anxiety free group", had the strong belief concerning their own performance seemed to sustain their confidence. "At first, I was scared about making a mistake, but I realized that I have to fill myself with positive feelings. I could make sure of my outcome. I should do my best, to get rid of negative feelings, and not care about something or somebody." After a while, some students in the "anxiety group" worried about things like: schedule, advertisement, music, audience, equipment, and also money. Furthermore, we may say that students who played the cast roles were included in the "anxiety free group", while students in stagehand roles were found to be in the "anxiety group". Finally, a factor analysis (varimax rotation) of each subject was conducted. The 3 factors were sufficient for the evaluation of their own feelings, and these factors labeled "metallic, pleasant, and concentrated" were identified (see Table 1).

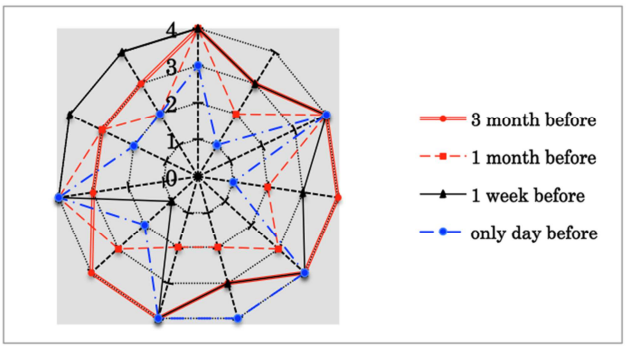
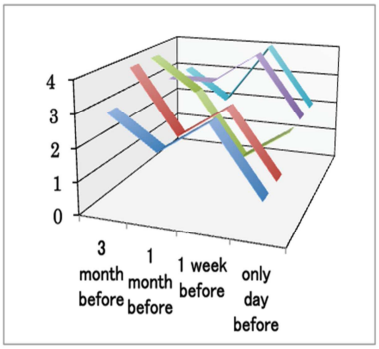


Figure 1. The transition of each subjects' anxiety across 4 investigations. Red double line =3 month before(1st), Red dotted line =1 month before(2nd), Black line = 1 week before(3rd), Blue dotted line =one day before(4th)



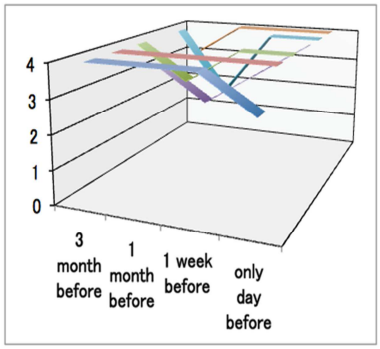


Figure 2. Anxiety free group **Figure 3. Anxiety group**

Table 1. Factor analysis of Study 1(varimax rotation)

	成分		
	1	2	3
quiet	.001	.265	.784
concentrated	-.123	.141	-.831
vitality	-.332	.597	.247
soft	-.633	-.057	.110
sharp	.518	.119	-.392
gentle	-.684	.065	-.309
nervous	.408	.296	.085
metallic	.810	-.181	-.127
pleasant	.102	-.835	.112
beautiful	.381	.304	-.069
clear	.303	.694	.124
distinct	.274	.803	.009
stable	.566	.018	.142
peaceful	.169	.150	.425
lucky	.186	-.603	-.496
shy	-.420	-.033	.662

3. STUDY 2

3.1. Method

Subjects. The participants were the same as in Study 1. The 11 university subjects, aged 20 to 24 years old participated.

Procedure. The procedure was similar to Study 1. All subjects were asked to judge their feelings about their regular concert performance on a STAI Y-1 sheet, and also fill in an SD sheet showing their emotion about their own performances. The entire test required approximately 15 minutes to complete.

3.2. Results and findings

The annual concert titled "The Graduation Concert 2012" was presented on the 16th of Feb. 2013. This concert was the final concert for the graduates, and allowed them to exhibit to an audience. In Figure 4, the blue dotted line showed a narrower area than other lines. Each movement of the 2 groups: anxiety group and anxiety free group was depicted in Figure 5 and 6.

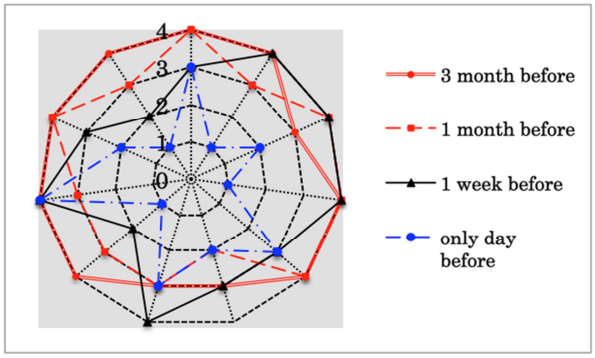


Figure 4. The transition of each subjects' anxiety across 4 investigations. Red double line =3 month before(1st), Red dotted line =1 month before(2nd), Black line = 1 week before(3rd), Blue dotted line =one day before(4th)

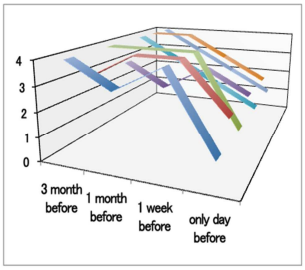


Figure 5. Anxiety free group

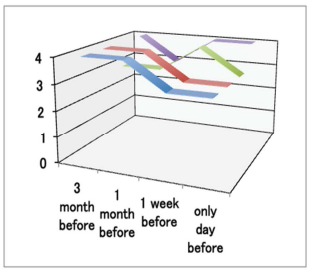


Figure 6. Anxiety group

To explain the difference of situation, a comparison between Figures 2, 3 and 5, 6 may be helpful. Figure 5 showed a more relaxed situation was prevalent among the annual concert. The number of students, without anxiety increased and the line between "1week before" and "one day before" fell sharper than in Figure 2. Furthermore, some students in the anxiety group showed lower anxiety levels than in Figure 3.

A factor analysis (varimax rotation) was conducted on each subject, as displayed in Table 2. The 3 factors labeled "clear, pleasant, and gentle" were extracted and identified for the evaluation of subjects' own feelings. Additionally, 2 kinds of SD profiles were shown in Figure 7 in order to compare two different situations. There were some similar curves on the adjective pair, such as "beautiful – ugly", "sharp – dull" and "stable – unstable". However, on the "quiet – noisy", "concentrated – non concentrated", "calm – nervous", and "pleasant – unpleasant" curves, a somewhat different pattern was found. We may indicate this through a curve highlighting which Opera was more unnerving for most university students, it was the annual concert.

Table 2. Factor analysis of Study 2(varimax rotation)

	成分		
	1	2	3
quiet	-.559	-.289	.474
concentrated	.030	.301	-.614
vitality	.646	-.188	.069
soft	-.701	.253	-.024
sharp	.065	.005	.468
gentle	.013	-.155	-.808
nervous	.077	-.680	.392
metallic	.320	.328	.241
pleasant	-.014	.869	-.113
beautiful	.620	-.145	.617
clear	.702	.412	-.413
distinct	.618	.268	.149
stable	.097	.160	.353
peaceful	.685	-.212	.132
lucky	.043	-.788	-.103
shy	-.213	.726	.281

When speaking about performance anxiety, the level of attitude has a big influence upon the performer's mental condition. Many students telling their enthusiasm of the annual concert have claimed that some kind of stress encouraged by oneself, and a different quality of stress was in existence. One student also said that she found stress a good way to use a positive mental attitude in rehearsal, for her piano performance. It must be noted that there were many different factors between the Opera and the annual concert. According to Wilson (2002), three sources of stress: trait anxiety, situational stress, and task mastery "vary independently", and "…whether anxiety is beneficial or detrimental to performance depends upon their interplay".

Our participants were fourth graders, specializing in music, and the performance using their own musical instrument was considerably familiar. It should also be added that most audiences were of friends and family. In the Opera, their experience was near to an amateur level and we had a variety of audience members: local people, the senior members of reunion, and kindergarten children. As I mentioned before, all students in stagehand roles had to contend with all sorts of tasks from handbill arrangements to ticket selling. Therefore, making the Opera has a more complicated process incorporating different stress in order to complete tasks. There is considerable validity with this concept, though it should not be pushed too far.

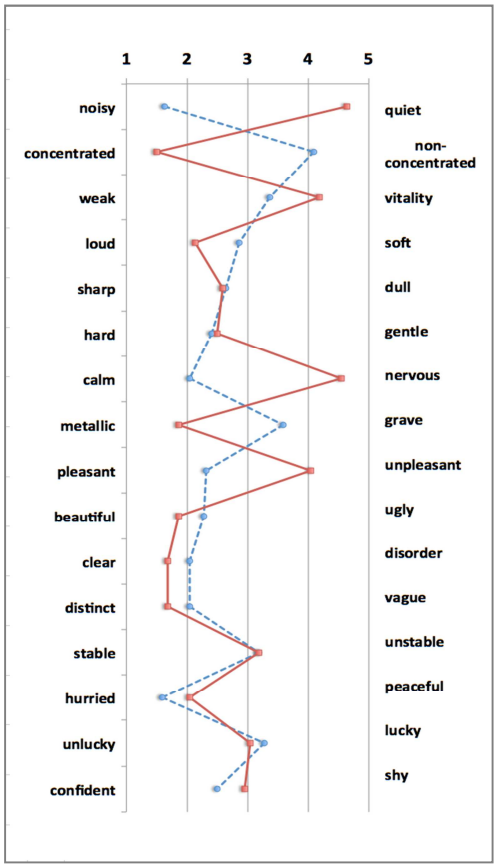


Figure 7. Mean ratings of SD profile in 2 conditions.
Red line =Opera, Blue dotted line =Annual concert

4. DISCUSSION

The present study was written with the aim of clarifying how university students react to performance anxiety under 2 different formal conditions: on Opera and the Annual concert. To follow is a summary of the results of the STAI Y-1 and the factor analysis of 16 pairs of adjectives as the following 7 items:

Opera

(1) Students were divided into 2 categories, anxiety free group and anxiety group.
(2) Some movements of subjects' anxiety tended to be reduced between the 3rd and 4th investigation.
(3) Students who played cast roles showed that cheerfulness overcame their anxiety.
(4) The 3 factors labeled "metallic, pleasant, and concentrated" were extracted.

Annual concert

(5) Students were also divided into 2 categories: anxiety free group and anxiety group, fewer students felt pressure.
(6) The degree of subjects' anxiety fell between the 3rd and 4th investigation.
(7) The 3 factors labeled "clear, pleasant, and gentle" were extracted and identified.

We are now ready to consider university students' performance anxiety about Opera and the annual concert. Although results suggest a similar uneasiness was found during the 3rd and 4th investigation in both situations, most students might find it easier to control their mental and physical state for their performance at the annual concert. Opera is so challenging, that students who help behind the scenes struggle to create a good image of the outcome. Meanwhile, many students played on the stage with plenty of confidence and roused their feeling by positive self-talk, irrespective of the 2 different conditions: Opera and the annual concert. It is widely accepted that solo performance is usually much more stressful than performing a duet or in a larger group, but there is no conclusive proof, established in this paper. It is debatable how students in stagehand roles control their anxiety. Of course, having these findings we still have a long way to go before we arrive at an interpretation of some of the differences existing around performance anxiety.

5. REFERENCES

Abel, J.L., & Larkin, K.T. (1990). Anticipation of performance among musicians: Physiological arousal, confidence and state anxiety. Psychology of Music, 18(2), *171-182.*

Steptoe, A. (1989). Stress, coping and stage fright in professional musicians. Psychology of Music, (17)1, *2-11.*

Wesner, R. B., Noyes, R., & Davis, T.L. (1990). The occurrence of performance anxiety among musicians. Journal of Affective Disorders, 18, *177-185.*

*Wilson, G.D., & Roland, D. (2002). Performance Anxiety. In Parncutt, R. & McPherson, G. E. (Eds.) T*he science & psychology of music performance. *Oxford, Oxford University Press.*

Wolfe, M. L. (1989). Correlates of adaptive and maladaptive musical performance anxiety. Medical Problems of Performing Artists, 4(1), *49.*

[1-18] CONTEXTUAL VARIABILITY IN AFFECTIVE RESPONSES TO MUSIC

Justin Black[1], Elizabeth Margulis[1]
[1]University of Arkansas, USA

Poster

1. Background: Much of music's power seems to come from its ability to affect our emotions. Currently there are a variety of theories that address the origin of emotional responses to music, ranging from perceived expectation all the way to situational based context. This research fails to address the relationship between musical and extramusical emotions. The ability for music to induce emotion in all situations is important in both scientific and real world settings.

2. Aims: The goal of this study is to induce an emotion and then looks at musical effects on it, to investigate music 's ability to induce strong enough emotions to overcome any previous emotional context.

3. Method: Several pieces of music were selected that, based on a pilot study, were shown to have a perceived emotional context. In order to create each of the desired emotional contexts, participants were randomly assigned to either write about an autobiographical event that was positive or negative, for ten minutes. Then participants were asked to listen to a song, which again depended on the assigned condition, either positive (happy) or negative (sad). Next participants were evaluated by performing the Positive and Negative Affect Schedule (PANAS-X) as well as an incentive rating task, in which participants rated how much they desired to perform a particular task.

4. Results: Results are expected to show that music is strong enough to override existing emotional context and induce the emotional characteristics found within the music, however, the strength of the results will vary across conditions. Negative emotions tend to be stronger and easier to induce than happy ones. Therefore, the extent to which positive musical stimuli influence or change emotional state when introduced into a negative context may be less intense than that of a negative musical stimuli in a positive context. Negative musical stimuli may even cathartically intensify negative emotions, given people's tendency to seek out sad music when they're feeling sad. Preliminary results also indicate a disconnect between participants' felt and their percieved emotional state

5. Conclusion: Music induction techniques are a valid option for inducing specific emotion in a variety of contexts.

[1-19] AROUSAL FACILITATES RECOGNITION MEMORY FOR UNFAMILIAR MUSIC

Madeline Huberth[1], Sarah Hawkins[2], Ian Cross[2]
[1]Center for Computer Research in Music and Acoustics, Stanford University, USA, [2]Centre for Music and Science, University of Cambridge, United Kingdom

Poster

The importance of emotional arousal when a stimulus is encoded has been demonstrated for many types of information. Generally, higher levels of arousal facilitate long-term memory. However, the few studies that investigated the arousal effect on memory for music yield contradictory and sometimes negative results, which merit further investigation. We tested the effect of arousal on memory using musical stimuli. We aimed to (1) determine if music that elicits high arousal is more likely to be subsequently recognized than pieces eliciting moderate or low arousal, and (2) examine the relative strength of arousal and valence as predictors of recognition. Eighteen participants' recognition of 16 musical excerpts from 19th-21st century symphonies (mean duration 47 s), screened to be unfamiliar, was tested over two sessions 24 hours apart. During each session, emotional response was evaluated by two 9-point Likert scales (arousal and valence) accompanied by self-assessment manikins. Response choices 'Yes', 'No', and 'Not sure' to 'Have you heard this piece before?' assessed recognition. Sixteen excerpts were presented in the first session and were randomized with 16 new excerpts during the second session. Skin conductance (EDA), using Q sensor technology, was measured during both sessions. Recognition scores rose significantly between sessions 1 and 2, and while arousal and valence ratings together influenced recognition, the effect was due entirely to arousal ratings, which significantly predicted increases in recognition. Arousal ratings correlated positively with recognition, but valence ratings negatively. EDA had no significant effect on recognition measures. These results show that more recognizable music is associated with increased arousal. Valence's negative correlation with high arousal may explain its negative correlation with recognition. EDA may not have supported the self-report findings because EDA increases with familiarity. To test this, future work could vary level of initial familiarity.

[1-20] PERCEIVED AFFECT OF MUSICAL INSTRUMENT TIMBRES

Chelsea Douglas[1], Stephen Mcadams[1]
[1]Cirmmt, Schulich School of Music, McGill University, Canada

Poster

1. Background: In addition to global factors of a piece, listeners use a combination of acoustic factors to assess emotional intent. As little as 250 milliseconds of a musical excerpt can hold enough information to perceive an emotional tone. The ability to recognize emotion in such short stimuli emphasizes the importance of examining how local musical factors, such as timbre, contribute to emotion perception.

2. Aims: This experiment facilitates a comparison between perceived affect and preference ratings of orchestral timbres, and examines the roles of register, attack, and playing technique.

3. Method: Recorded instrumental samples were chosen from the Vienna Symphonic Library. Stimuli had a constant duration (500 ms), pitch class (D#), and dynamic (forte). Samples were chosen from the entire range of each instrument within D#1 to D#8 with different attacks (weak, normal, or strong). Techniques like flutter-tonguing for brass and woodwinds and pizzicato for strings were included. Participants completed six ratings per sound on 9-point bipolar scales. Four scales measured perceived emotion (valence: negative/positive; valence: displeasure/pleasure;

energy arousal: tired/awake; tension arousal: tense/relaxed) and two measured participants' preference (dislike/like) and familiarity (unfamiliar/familiar).

4. Results: There was a strong, positive correlation between preference and valence ratings, and a moderately strong negative correlation between preference and tension ratings. There was only a weak correlation between preference and energy ratings. Energy ratings were positively correlated with register: timbres in higher registers were perceived as more awake. String and percussion timbres tended to have higher valence and energy ratings than brass and woodwind timbres.

5. Conclusion: Despite the short duration of the stimuli, participants assigned ratings representative of a three-dimensional affect scale to the various timbres. The dimensions of perceived valence and preference were positively correlated, but not identical (a perceived negative valence was not necessarily associate with a feeling of dislike), and were thus analyzed separately. It is likely that spectral descriptors of timbre are closely related to perceived energy ratings. Further analysis of the timbres may provide information to further map specific spectral or temporal properties onto the perceived affect ratings. Timbre should be considered as a prominent vehicle of affective expression in music.

[1-21] MUSICAL FEATURES OF SPONTANEOUS IMPROVISATION ASSOCIATED WITH EMOTIONAL CUES

Malinda Mcpherson[1], Monica Lopez-Gonzalez[1], Summer Rankin[1], Charles Limb[1]
[1]Department of Otolaryngology - Head and Neck Surgery, Johns Hopkins University School of Medicine, USA
Poster

1. Background: A primary aspect of music as a form of auditory communication is its ability to convey emotion. Yet how music achieves this remarkable capacity to both express and induce emotion remains unclear. While each component of music (e.g. key, mode, tempo, etc.) contributes to the ability of music to convey emotion, no single feature of music sufficiently accounts for the vast emotional range of music. Attempts to characterize the musical expression of emotions have examined this issue using simplified experimental methods that fail to consider the breadth of musical expression.

2. Aims: To study the links between musical content and expression using a more musically and ecologically valid method, we examined several features of musical improvisations generated in response to an emotional target. We hypothesized that this approach would allow us to identify the distribution of musical elements utilized to convey emotion.

3. Method: Fourteen professional jazz pianists were asked to improvise compositions in response to happy, sad and ambiguous emotional targets, presented using photographs of faces and still cartoon faces. There were no musical constraints other than duration (improvisations were sixty seconds). All improvisations were recorded using MIDI and analyzed in MATLAB to determine their mode, key, range, note density (tempo), average note duration, volume, and articulation (legato vs. staccato).

4. Results: Although there were significant differences for each musical measure between emotions, significant overlap was also observed. In general, happy improvisations were more likely to be in major keys, have faster tempos, faster key press velocities and more staccato notes when compared to sad improvisations, with ambiguous improvisations generally falling closer to sad improvisations. There was a wide distribution and many outliers for every musical feature.

5. Conclusions: While we found notable differences between the happy and sad improvisations, the wide variety of musical content improvised for each emotion was equally striking. Our study supports the claim that the expression of emotion through music cannot be fully described using simplified models of musical features. Instead, our results reveal that while specific musical variables are used to express specific emotions, the range of features used is far wider than previously shown.

[1-22] PSYCHOLOGICAL EFFECTS OF 'HITOKARA' SINGING ON MOOD

Junko Matsumoto[1]
[1]Nagano College of Nursing, Japan
Poster

Previous studies have reported that singing karaoke is associated with positive psychological and physiological effects. When people sing karaoke, generally more than one person participates in the session. However, recently a trend of one parson going alone to karaoke has been seen. This is called "hitokara" (solo karaoke) and there are even special hitokara karaoke booths. In this study we investigated impressions of hitokara, reasons why people go to hitokara, and mood induced by hitokara. One hundred ninety-two college students completed a questionnaire reporting on their usual participation in karaoke and also their participation in hitokara. The results suggest that hitokara is popular among college students, but that not many students had done hitokara. The students felt that hitokara allows one to sing many songs alone and at their own pace. Those with at least one hitokara experience felt refreshed after singing hitokara, and the moods induced after hitokara were similar to those induced after usual karaoke. However, mood intensity after hitokara was lower than that after usual karaoke. Consequently, the psychological effects of hitokara may be weaker than those of usual karaoke, but it does have a beneficial effect on mood.

Poster Session 2

Baekyang Building Lobby
16:15 – 17:00, Tuesday 5 Aug 2014

[2-01] IMPROVISATION AND CHANGE IN VIDEOS OF 1-TO-1 MUSIC THERAPY SESSIONS WITH CHILDREN WITH AUTISM SPECTRUM DISORDERS: A CASE EXAMPLE

Neta Spiro[1], Chloe Rush[2], Tommi Himberg[3]
[1]Research Department, Nordoff Robbins, United Kingdom, [2]Centre for Music and Science, Faculty of Music, University of Cambridge, United Kingdom, [3]Brain Research Unit, O.v. Lounasmaa Laboratory, Aalto University, Finland
Poster

ABSTRACT

The individual and shared pulse characteristics of participants in interactive and co-improvisational music therapy approaches are often described as one of the reasons that music therapy has been found to be effective for clients with autism spectrum disorders. Music therapy works towards change but the documentation and analysis of change varies depending on the music therapy approach and the purpose of the analysis. In this case example, we analyse videos of one early and one later Nordoff Robbins music therapy session using an annotation protocol in order to investigate pulse characteristics of both players and to examine whether change can be identified in the individual player's pulse profile and in the amount of shared pulse. We find that instances of shared pulse primarily occurred within a regular pulse, and more regular and shared pulse behaviors were noted in the later session. Pulse characteristics may be taken as an indicator of client-therapist interaction and form part of a web of characteristics that our investigations of improvisation, interaction and change in music therapy sessions explore.

1. INTRODUCTION

Music therapy has been shown to help children with autism spectrum disorders (ASD) but the processes that occur during music therapy still require investigation (Gold et al., 2006). Nordoff-Robbins – as a music-centred approach to music therapy – is an improvised, interactive approach in which musical and personal change are seen as interlinked (Aigen, 2005 Brown, 1994, Nordoff & Robbins, 2007). During Nordoff-Robbins music therapy sessions the therapist monitors and regulates their own music making according to the client's music, gestures, vocalizing and movements, often with the initial aim of establishing, sustaining and developing a shared, flexible and negotiated regular pulse (Bruscia, 1987). In Nordoff and Robbins' conception, the "basic beat" is understood to be the basis for developing a mutually negotiated musical relationship in music therapy (Nordoff and Robbins, 2007, p. 298, see also: Aigen, 2005, Ansdell 1996, Pavlicevic 2000, Wigram, 2004).

Studies in music perception and production have suggested that typically developing children entrain their pulse to another pulse from a young age and are better at doing so with other people than with automatons (e.g. Kirschner and Tomasello, 2009). This does not seem to happen so readily for at least some children with ASD diagnoses and, though the studies available on the topic are limited and reasons for difficulty not clear, it seems that rhythmic co-ordination can be problematic for children with ASD (e.g. Isenhower et al., 2012). Such difficulties of self-regulation of one's own timing and synchrony with others have been suggested as important in the understanding social experiences of individuals with autism (e.g. Hardy and LaGasse, 2013). Taken together with studies suggesting that entrainment can contribute to social affiliation (e.g. Hove and Risen, 2009), a detailed investigation of pulse characteristics of interaction in music therapy may provide insight into how such characteristics change over time in music therapy and, eventually the effect they have for clients beyond the therapy room.

Improvisational music therapy with children with ASD provides an opportunity to explore the characteristics of a 'special' kind of interaction; often with little verbal interaction, the same people return frequently, regularly over several years, and there is explicit focus on aspects of communication and development. Explorations of such contexts therefore have potential to teach us about interaction and communication in music more generally.

2. AIM

The aim of this part of the study is to identify patterns of change in basic characteristics of 1-to-1 music therapy sessions with children with ASD. Specifically, we focus on the changing pulse relationship between the players.

For clients with autism, stability in the structure of their environment is seen as crucial (Wigram and Gold, 2006). In music therapy, part of the role of the therapist is to provide 'rhythmic grounding' (Bruscia, 1987; Wigram, 2004), creating order out of 'chaos' to offer regular temporal structures within the improvised musical environment. This is one reason why music therapy has been seen to be superior to forms of intervention in which music was not used (Gold, et al., 2006).

In a previous case study that used the same basic methods as employed here, the pulse behavior of the therapist was found to be predominantly regular, and the majority of the client's regular playing was found to occur at the same time as the therapist's regular playing (Spiro et al., 2013). That was a study of one client-therapist pair in their first and second music therapy sessions.

In the current study, we investigate a longer period of intervention asking the question 'Do individual pulse behaviors of client and therapist change over the course of intervention?' Predictability of pulse may be a first step towards synchrony and so we also ask: Do bouts of regular pulse behavior coincide with shared pulse? Players can manifest their shared pulse by playing at the same time as each other, or to continue each other's pulse when turn taking and we explore which of these behaviors occurs. Finally, we explore to what extent there are differences between the early and later sessions in individual and relative pulse characteristics of the players.

3. METHODS

Participants

The client, diagnosed with ASD is a boy, aged 4 in the early session (born October 2003). The music therapist is female, trained at Nordoff Robbins.

Materials

Two videos, 26 min 36 s and 27 min 57 s in duration, were analysed. Both were filmed at the Nordoff Robbins London Centre. The *early* session is the initial music therapy session for this client-therapist pair, filmed on 26/05/2007. The *later* session features the same pair approximately one year later and was filmed on 28/06/2008. The client and therapist saw each other more or less weekly between these sessions. Ethical approval was granted by the NR and the Faculty of Music, University of Cambridge Research Ethics Committees.

The videos used in the current study are those collected during the usual clinical work of the clients and therapists. Therefore, we have little control of recording methods or quality. The videos were made by switching between two CCTV cameras and with one microphone for sound. Given the nature of the recordings, automatic analysis of the video and audio was not practical. Therefore, manual annotation based on what can be seen and heard is used.

Annotation Protocol. Here we report on the results of a subset of a total of 27 tiers annotated (Rush, 2013). The annotation protocol for the pulse tiers is presented in Figure 1. Once annotations of either regular, irregular or no pulse had been assigned to the entire music therapy session, moments in which client and therapist were perceived to be playing the same pulse simultaneously were marked in a separate tier: shared pulse. Where musical articulations were judged to be in the same a metrical framework, they were marked as 'shared pulse'.

During moments of ambiguity, body movements were taken into account. This is similar to the method used by Spiro et al. (2013) but also includes the possibility for moments of irregular shared pulse.

Periods of time identified as having a *shared pulse* were then labeled as either moments of *synchronization* or *turn taking*. *Synchronization* was annotated for moments during which the client and therapist made rhythmic musical articulations at same time. *Turn taking* was annotated when the client and therapist continued each other's pulse one after another.

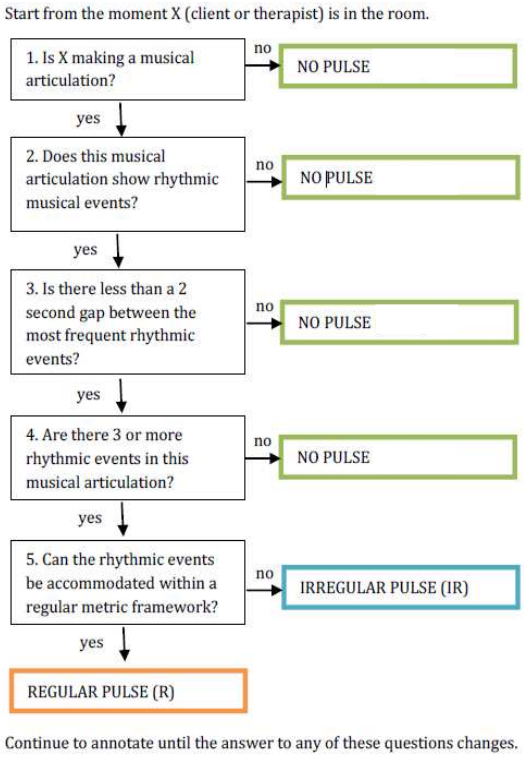


Figure 1: Excerpt of Annotation protocol: Pulse related annotations.

Definitions of key terms:

(1) A **musical articulation** is defined as any musical gesture, usually resulting in a sound, produced intentionally by the individual (excludes typical conversational speech and includes body movement when music is present).

(2) **Rhythmic musical events** are accented musical articulations that give rise to a rhythmic profile (irregular or regular): this excludes musical articulations that do not have a clear rhythmic profile, such as a drum roll or a piano tremolo, and that are not accompanied by rhythmic bodily gestures from the creator.

(3) The **two second time gap** between rhythmic events is in a place as this is understood to be the upper limit for entrainment (Handel, 1987).

(4) **Three of more rhythmic events** are required in order to subjectively gauge the relationship between the inter-onset intervals and make a pulse judgment. Therefore, a pulse judgment could not be made from two isolated piano chords (Spiro et al., 2013).

Manual annotation in ELAN (http://tla.mpi.nl/tools/tla-tools/elan/) was followed by analysis of proportions of instances and duration and the relationship between annotations of clients' and therapists' behaviors. In addition, time-line visualizations of annotations were produced in MATLAB using custom functions and the SALEM toolbox (http://aiweb.techfak.uni-bielefeld.de/c/elan-matlab).

Inter-annotator reliability. A co-annotator analyzed 10% of the video material which had been annotated by the second author using the same guidelines. Divergent annotations were noted and statistical analysis tests for correlation between annotators were applied. Pearson's Rank was used to test for correlation between durations of moments identified by annotator and co-annotator. Cohen's Kappa was used to test for consistency between annotators in labeling annotations that occurred at the same time.

4. RESULTS

4.1. Inter-annotator reliability

Sixty-eight percent of all annotations made by annotator and co-annotator were identical, and the correlation results show that there was a significant positive relationship between annotators for annotation duration (early video, $r = 0.77$, $p < 0.001$; later video, $r = 0.80$, $p < 0.001$). There was also substantial agreement for labels assigned to matching annotations (early session, $K = 0.65$, $p < 0.001$; late session, $K = 0.66$, $p < 0.001$), indicating that annotations made by both annotators tended to follow the same sequence, with only small displacements in time.

More specifically, a measure for the agreement of the start and end times of all pulse-related annotations was calculated. Results are displayed in Table 1. The correlation of durations of annotations made by annotator and co-annotator was also calculated and this indicated a strong, positive relationship. The data were thus deemed reliable for use in further analysis.

Table 1: Column 1: Pearson's correlation values indicating agreement between annotator and co-annotator durations of annotations. Columns 2 & 3: Average difference between annotator and co-annotator start and end times as proportions of the average annotation duration for pulse tiers.

	Duration (s) correlation	Average difference (%): start times	Average difference (%): end times
Client pulse	$r = 0.702$, $p = 0.052$	14.9	32.6
Therapist pulse	$r = 0.997$, $p < 0.01$	8.7	6.6

For bouts of client pulse, the average annotation duration was 4.1 s, and for bouts of therapist pulse, the average duration was 11.3 s. This explains the larger variability indicated by the results of the client-pulse tier, where differences could span up to one third of the average annotation length.

4.2. Pulse characteristics and indicators of interaction

The client spent 48% of the early session and 38% of the later session making sounds while the therapist spent 69% and 68% of the two sessions making sound, respectively.

The overview of annotated pulse behaviors in Figure 2 shows that regular pulse occurred more frequently in the therapist's behavior than in the client's, and the client made more sounds with pulse (either regular or irregular) in the late session than early one. The difference was larger for the regular than irregular pulse.

Figure 3 depicts the patterns and co-occurrence of behaviors of the two players across the two sessions on a time-line, with each annotation tier represented as a row on the graph. The two graphs in Figure 3 show the annotations of the individual and relative pulse characteristics of the early (a) and later (b) videos. (The graphs show only the durations annotated as regular and irregular pulse and exclude the durations for which there was not sound or no pulse.)

Viewed in this way we see that, broadly speaking, both the client and the therapist engage in more sustained rhythmic production in the later session. This is indicated by the more continuous blocks of annotation. Moreover, even in the early session, the longer blocks of activity are mainly those in which the music therapist or client (or both) had regular pulse and there are more of these in the later session than in the earlier one.

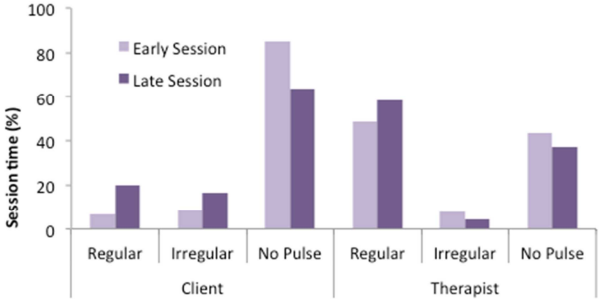


Figure 2: Total proportion of time annotated with each pulse behavior for both players in both sessions

Looking at the third tier in each graph, there are more instances of shared behavior in the late than the early session and most of the client's regular playing occurred at the same time as the therapists' regular playing. The average duration for a shared pulse occurrence was 7.7 s in the early video (SD 11.6 s) compared with 25.3 s in the later video (SD 25.7 s) ($t(20) = 1.901$, $p < 0.05$). Shared pulse bouts were the longest bouts the client produced.

Comparing these three tiers in each session, the majority of shared pulse behaviors (shared pulse synchrony or turn taking) occurred in moments of shared regular pulse rather than shared irregular pulse. The longest bout of shared pulse was in the form of turn taking and occurred in the later session, while there was almost no turn taking in the first.

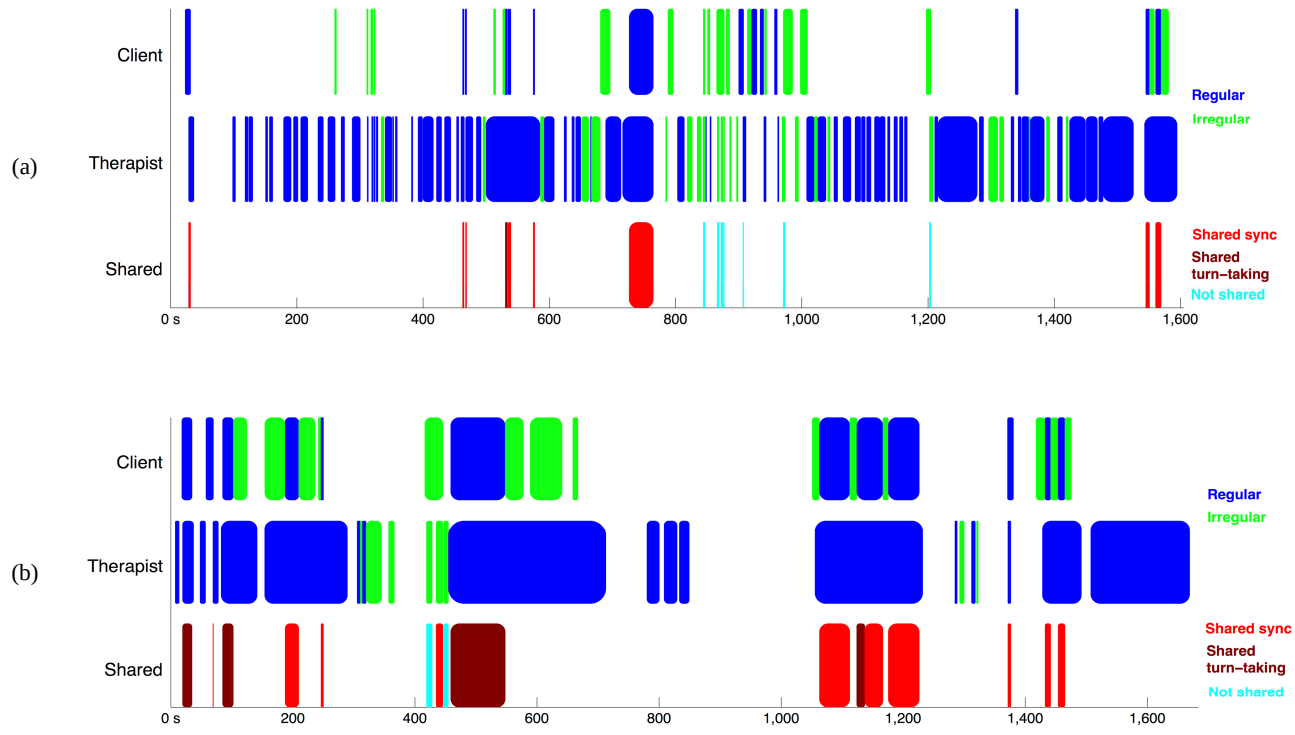


Figure 3: Pulse characteristics throughout the (a) early and (b) later videos

5. DISCUSSION

In this case example, comparison of two music therapy sessions one year apart revealed differences in pulse behaviors of the client-therapist pair across sessions. The client engaged in more sustained pulse production in the later session, including longer bouts of synchrony and turn taking. There was a greater co-occurrence of shared pulse within a regular pulse framework, and a greater number of instances of these behaviors were noted after one year of intervention. In both sessions, the client produced a regular pulse for a much smaller proportion of the time than the therapist. These findings are promising given what they may indicate regarding changes in the client's temporal self-regulation and interaction with the therapist.

This analysis was based on identifying bouts of shared pulse by listening to and watching the videos. A more detailed analysis of the articulation timings would be necessary for characterizing the client-therapist interaction in more detail, e.g. by indicating whether the synchrony was a result of co-adaptation (entrainment) or unidirectional adaptation. Furthermore, a fuller analysis of the data set will include the music therapist's interpretation of these sessions.

Using video analysis, a subset of basic characteristics of behavior of the client and therapist have been analyzed using a replicable annotation protocol, making use of succinct measures that can be used as indicators of change which are intended to be applicable across different sessions, clients and therapists.

6. ACKNOWLEDGEMENTS

Many thanks to the music therapist and client at Nordoff Robbins' London Centre for generously providing videos, to Francesca Rogers for co-annotating the videos and to members of the Nordoff Robbins Research Department, in particular Camilla Farrant and Mércèdes Pavlicevic.

7. REFERENCES

Aigen, K. (2005). Music-centered Music Therapy. *Gilsum NH: Barcelona Publishers.*

Ansdell, G. (1996). Talking about Music Therapy: A Dilemma and a Qualitative Experiment. British Journal of Music Therapy, *10, 4–16*

Bruscia, K. E. (1987). Improvisational Models of Music Therapy. Springfield, IL: Charles C Thomas Publications.

Brown, S. (1994). Autism and music therapy – Is change possible, and why music? British Journal of Music Therapy, 8*(1), 15–25.*

Gold, C., Wigram, T., & Elefant, C. (2006). Music therapy for autistic spectrum disorder. Cochrane Database Syst Rev, 2.

Handel, S. (1987). Listening: an Introduction to the Perception of Auditory Events. *MIT Press.*

Hardy M. W. & LaGasse A. B. (2013). Rhythm, movement, and autism: using rhythmic rehabilitation research as a model for autism. Front. Integr. Neurosci. 7*(19).*

Hove, M. J. & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. Social Cognition, 27*(6), 949–960.*

Isenhower, R.W, Marsh, K.L., Richardson, M.J., Helt M., Schmidt, R.C. & Fein, D. (2012). Rhythmic bimanual coordination is impaired in young children with autism spectrum disorder. Research in Autism Spectrum Disorders 6, *25–31.*

Pavlicevic, M. (2000). Improvisation in Music Therapy: Human Communication in Sound. Journal of Music Therapy, 37*(4), 269–285*

Nordoff, P. & Robbins, C. (2007). Creative Music Therapy: A Guide to Fostering Clinical Musicianship (2nd Edition). *Gilsum NH: Barcelona Publishers.*

Rush, C. (2013) Timing in music therapy: an investigation of relative pulse behaviours of a child with autism and his therapist. *Unpublished Thesis.*

*Spiro, N., Schofield, M., & Himberg, T. (2013). Empathy in Musical Interaction. In P*roceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th–15th June 2013. Geoff Luck & Olivier Brabant (Eds.). University of Jyväskylä, Department of Music.

Wigram, T. (2004). Improvisation: Methods and Techniques for Music Therapy Clinicians, Educators and Students. *London and Philadelphia: Jessica Kingsley Publishers.*

Wigram, T. & Gold, C. (2006). Music therapy in the assessment and treatment of autistic spectrum disorder: clinical application and research evidence. Child: care, health and development, 32*(5), 535–542.*

[2-02] USING MUSIC VISUALIZATION SYSTEMS WITH HEARING-IMPAIRED CHILDREN

Rumi Hiraga[1], Kazuhiko Otsuka[1], Shu Matsuda[2], Keitaro Takahashi[3]
[1]Tsukuba University of Technology, Japan, [2]Paox Inc., Japan, [3]Basel Musik-akademie, Switzerland
Poster

ABSTRACT

With our more than ten years of experience working with hearing-impaired college students, we know that many of these individuals enjoy listening to music. There have been reports on the positive effect of musical activities on hearing-impaired children when it comes to learning language, and we believe that providing good musical experiences for such students will be effective in terms of receptivity of sound. We have developed two visualization systems in pursuit of this goal and visited several schools and special education classes for deaf and hard of hearing children to let them use the systems. They were very excited about using the systems and indicated that they appreciated them. Comments from teachers and students have inspired us to continue developing our systems to meet their expectations.

1. INTRODUCTION

Our more than ten years of experience working with hearing-impaired college students has made it clear to us that they enjoy music very much. They not only listen to music but also sing songs at Karaoke, play music games such as Dance Dance Revolution and Drum Master, which were arcade games in the past but now commonly played on PCs and tablets, dance to music, and even play real musical instruments. We have interviewed students about their musical activities 0, and know that deaf and hard of hearing students have had musical experiences in their childhood education regardless of whether they went to special schools for hearing disabilities or to regular schools. Some students had received private instruction on the piano. When working on their graduation theses at the Tsukuba University of Technology, a national university in Japan for hearing-impaired students (and also visually impaired students in a different campus), about a quarter of students selected a music-related theme, such as "a music game for hearing-impaired people" and "a karaoke practice system."

The interest of hearing-impaired students in sound raises their speculation on the perception of sounds. For example, they notice that those who have the same hearing level differ in listening comprehension in spoken language. They also feel uncertainty in listening to music—is their perception of a song the same as or different from that of a hearing person? If the reason for the different understanding of sounds with the same hearing level and the difference between what hearing people and hard of hearing people listen to in music can be identified, then the relationship between hearing-impaired people and sound will change and the results will enhance their quality of life. These are interesting research issues to investigate.

Our goal in this work is to provide hearing-impaired children with music via computer software systems, thus giving them musical experiences they have not had the opportunity to explore before. We hope that eventually children will be able to enjoy music spontaneously and continuously with the experiments with the systems. In this paper, we describe two visualization systems that we developed and discuss how hearing-impaired children interacted with them. After introducing the systems to hearing-impaired children, we received helpful feedback from both teachers and students that we are now taking into consideration as we improve our systems to more adequately meet the users' interest.

2. RELATED WORKS

Music is not necessarily impossible for deaf and hard of hearing people to enjoy. There are even professional musicians active today, including Dame Evelyn Glennie, a percussionist, and Paul Whittaker OBE, who has been deaf since birth and plays the piano and organ as well as trains choirs. The Gallaudet Dance Company, founded in 1955, and Chinese Disabled People's Performing Art Troupe give performances all over the world.

Although not all hearing-impaired people have opportunities to enjoy music, there is some work being done by various musicians and therapists to provide music for hearing-impaired people. A-A Darrow facilitated several music therapy sessions to hearing-impaired children 0. Pioneer, a Japanese electric company, holds regular concerts that enable hearing-impaired people to enjoy music by providing body sonic chairs and information support with captions and sign language translation. K. Sato, a music composer, is engaged in various activities to help hearing-impaired people enjoy music. Hiraga offered a computer music class to hearing-impaired college students in which students played Miburi, a wearable MIDI instrument developed by YAMAHA, in their own playing method that YAMAHA did not imagine 0.

The recognition and effect of music by and to hearing-impaired people has been studied by several research groups. Hiraga reported there were no differences between deaf/hard of hearing people and hearing people in recognizing certain emotions expressed in drum performances 0. They also reported that effect of accompanying visual information acts as a cue for recognizing an emotion in music 0. However, there were significant differences between the two groups in the recognition of timbre 0. Other research groups have found that children who have cochlear implants and perform musical activities have a better understanding of speech. Torppa et al. found that musical experiences have a strong link to the perception of focus in children with cochlear implants 0, while Trehub et al. found that such children could recognize music even when the stimuli had been altered 0.

3. VISUALIZATION SYSTEMS

We developed two types of visualization systems. The first one, which corresponds to musical performances with MIDI instruments, is called Walking, and the second one, in which sound input is used to move a still image, is called Moving Picture with Music (MPM).



(a) Through fog. WP is walking on the balloons.



(b) Over grass



(c) Within view of Mt. Fuji.
WP almost jumps out of the scene.

Figure 1: Screen shots of Walking

3.1. Walking

With Walking, users can see image moving according to their manipulation of MIDI instruments. On the Walking screen (Fig. 1), a person (hereafter we call the person Walking Person, WP) walks through fog, over the grass, within view of Mt. Fuji, and over the surface of water. These background scenes change automatically and move from right to left as if WP in the scene steps from left to right.

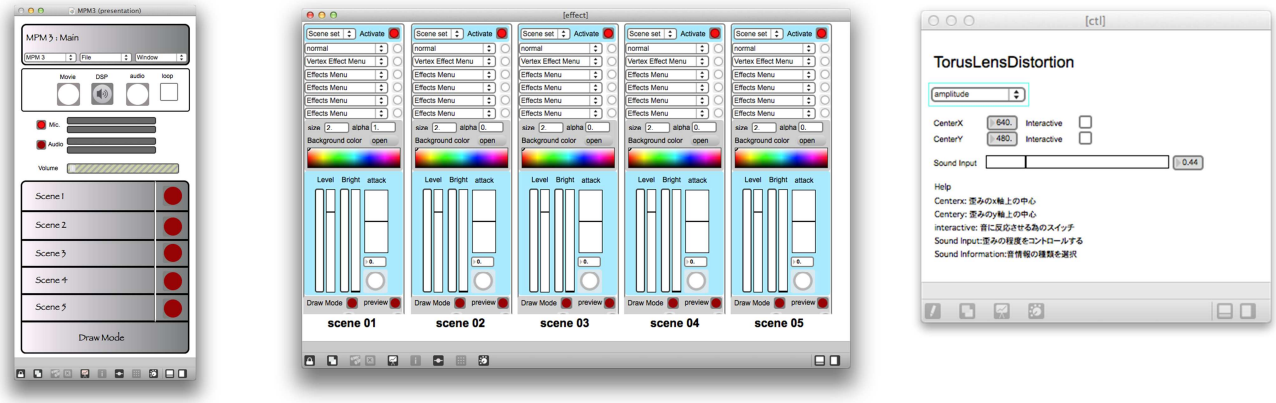
Two MIDI devices are used to manipulate the image: one to generate balloons and the other to let WP change his/her vertical position as if he/her jumps. The first MIDI instrument determines the color and size of a balloon by the MIDI note number and the velocity, respectively. The vertical position of the balloons is ambiguous. With the input from the second MIDI device, WP jumps up as in Fig. 1 (a) and (c). The vertical positions of balloons and of WP are determined independently.

We programmed Walking with Max/MSP/Jitter and Digital Image Processing with Sound (DIPS) 0. DIPS provides a set of Max objects that handle OpenGL functions, OpenGL Shading Language (GLSL), and Apple Core Image, separately from real-time visual-image processing events.

3.2. Moving Picture with Music (MPM)

With MPM, users choose a still image, plan how to move it, and then make a sound to move the image. MPM also runs on Max/MSP/Jitter with DIPS. MPM implements sets of objects for effects and for generating video files. With MPM, Hiraga generated moving images from a still image that itself does not recall any emotions for the purpose of recognizing emotions in movement 0.

Figure 2 (a) shows the main window of MPM. The sound input is either through a microphone or the replay of a WAV audio file. A scene is a set consisting of a still image and a way to move. In MPM, users can set up to five scenes. If they preset scenes, they can change the images and the movements during the music performance. Figure 2 (b) shows the effect window. Users choose a still image and then open a control window to decide how to move the image. Figure 2 (c) shows a type of control window. Users can choose the degree of effect and sets parameter values.



(a) Main Window (b) Effect Window (c) Control Window

Figure 2: Windows of MPM

Table 1 shows some of the effects that MPM provides. Control windows are prepared for each of the effects. The effects can be accumulated, allowing users to come up with their own unique movement style.

Figure 3 shows example shots of moving images. When a user chooses a still image (Fig. 3 (a)), it moves according to the type of effect. Fig. 3 (b) and (c) show screen shots when the image is moved with the vertex effect. The torus lens distortion and motion blur effects are shown in Fig. 3 (d) and (e), respectively

Table 1: MPM movement effects.

Effect	Description
Vertex effect	Using the noise functions of GLSL. Parameters are speed, density, and strength.
Distortions	Including bump, twirl, torus lens, pinch, and hole distortions.
Mosaic	Both a regular and random mosaic.
Blur	Including motion, recursive, Gaussian, radial and zoom blurs.
Others	Repeater, edges, and gamma adjust.

4. USING SYSTEMS

We visited four classes (three elementary school classes and a kindergarten class) and showed our systems to a total of nine kindergarten kids, twenty-seven elementary school students, and six junior high school students. The visits were as follows:

(1) Saga School for the Deaf. Nine elementary school students and six junior high school students.

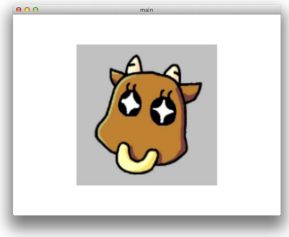
(2) Kasumigaura School for the Deaf. Nine kindergarten kids and fourteen elementary school students.

(3) Takezono-Higashi (TH) special support class. Four elementary school students.

Saga and Kasumigaura are provincial cities, and TH is located in the center of Tsukuba Science City. Scientists often visit TH, and the children there are accustomed to handling PCs and tablets.

In addition to a MacBook Pro, on which our two systems were installed, we brought two speakers (audio-technica, Boogie Box AT-SPB5), synthesizers with a keyboard (micro KORG and Pianist Pro, an iPad application), and several percussions (a guiro and others). A new type of instrument, “Tenori-on,” did not interest the students much. The school we visited had music rooms equipped with MIDI keyboards and other instruments including a Japanese Taiko drum, which is often used in music class for hearing-impaired children.

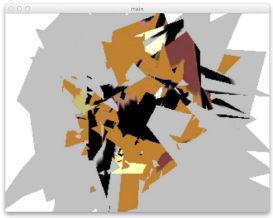
In all three schools, the children enjoyed using the systems very much, and the teachers were pleased with the students’ strong interest in using the music instruments and in manipulating the images to make them move. The TH students also used the MPM effect and control windows—in other words, they not only played instruments and made sounds, they also made effects.



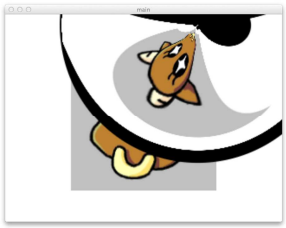
(a) Original still image.



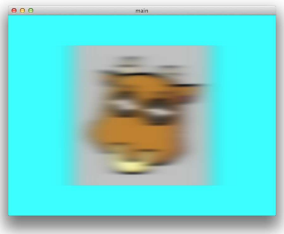
(b) Movement with vertex noise effect



(c) Movement with vertex noise effect



(d) Movement with torus lens effect



(e) Movement with motion blur effect.

Figure 3: MPM screen shots

4.1. Comments

After the visits, students and teachers provided us with comments on using systems. The following are a few excerpts. The systems should take into account hearing ability and school age. For those who cannot perceive sound, it is difficult to understand the varieties in the appearance of balloons. A clearer correspondence between an action performed by striking a key on the keyboard and the color, position, and size of the visualized object (balloon) is required. It is wonderful to share a mood with an image that is drawn from a musical performance. A story can be expressed with the image—for example, a chick hatches from a still image of an egg with the sound of a fowl.

4.2. Discussion

In general, the kids treated our systems as if the sole function was to play games, while their teachers wanted the students to know that they were controlling their vocalization with the visual images. Along with the comments that were provided, we can classify the issues as follows. Designing the systems: With Walking, the relationship between cause (sound) and effect (visual object or its move) should be clarified. Teachers want students to understand that playing the MIDI keyboard, i.e. playing music, is different from controlling other input devices. When some students used Walking, they could control the vertical position of WP in the scene as if WP was walking on balloons. With this type of operation, it is better if two students play music: one's performance shows balloons and the other's makes WP walk by playing music. It turned out that it was difficult for many of the students to make effects in MPM. Providing a hardware device to change effects is one way we could help children change the moving effects by themselves more easily. Using the systems: We should collaborate with teachers to find out how to best use music visualization systems that encourage students to maintain an interest in music and sound. Both in Walking and MPM, the scene can be shown to make a story—in Walking, there can be characters and background scenes that match the story, and in MPM, multi-effect function can change the scene in a story. As for sharing a mood, using MPM to play "body percussion" with a group of students can help them this feeling of sharing. A still image moves quietly or slowly if they all play softly, and the image moves fiercely if some of the students play strongly.

5. ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers 23611005 and 26282001.

6. REFERENCES

M. Matsubara, H. Terasawa, K. F. Hansen, and R. Hiraga. (2014). An inquiry into hearing-impaired student's musical activities – How do they listen to music?, ICMPC 2014 (to appear).

A.-A. Darrow (2006). The role of music in deaf culture: Deaf students' perception of emotion in music, Journal of Music Therapy, XLIII (1):2-15.

R. Hiraga and M. Kawashima (2004). Performance Visualization for Hearing Impaired Students. International Conference on Education and Information Systems: Technologies and Applications, Vol. 1, 323-328.
R. Hiraga and N. Kato (2006). Understanding emotion through multimedia—comparison between hearing-impaired people and people with hearing abilities. ACM International Conference on Computers and Accessibility, 141-148.
R. Hiraga, N. Kato, and N. Matsuda (2008). Effect of visual representation in recognizing emotion expressed in a musical performance. IEEE International Conference on Systems, Man, and Cybernetics, 131-136.
R. Hiraga and K. Otsuka (2012). On the recognition of Timbre, A first step toward understanding how hearing-impaired people perceive timbre. IEEE International Conference on Systems, Man, and Cybernetics, 2013-2018.
R. Torppa, A. Faulkner, J. Järvikivi, and M. Vainio (2010). Acquisition of focus by normal hearing and cochlear implanted children: The role of musical experience. International Conference on Speech Prosody.
S. E. Trehub, T. Vongpaisal, and T. Nakata (2009). Music in the lives of deaf children with cochlear implants. The Neurosciences and Music III: Disorders and Plasticity, Volume 1169.
DIPS, http://en.wikipedia.org/wiki/DIPS_%28Digital_Image_Processing_with_Sound%29
R. Hiraga and K. Takahashi (2011). Conveying Emotion with Moving Images: relationship between movement and emotion. Affective Computing and Intelligent Interaction, LNCS 6974, 558-567.

[2-03] WELL-BEING AND QUALITY OF LIFE: SELF-REFLECTION OF ADULT AMATEUR INSTRUMENTAL STUDENTS

Jesper Hohagen[1]
[1]*University of Hamburg, Germany*

Poster

1. Background: It is well-known that the reception and practice of musical activities promote physiological, social and mental well-being (MacDonald et al., 2012). Different surveys prove positive effects on health and sustainment by the attendance of cultural events (e.g. Johansson, 2001) and the strong influence of music on quality of life is evident (Ruud, 1997). Making music, listening and singing can enhance the physiological state of human body (e.g. Kreutz et al., 2004). Furthermore it is explored precisely that singing in a group or in a choir promotes health and well-being (e.g. Clift & Hancox, 2001). However, studies about the benefits of playing instruments on perception of quality of life and well-being are rarer.

2. Aims: The present survey investigates to what extent adult participants of amateur instrumental lessons feel benefits of their physiological, social, and emotional well-being and their quality of life by their regular musical activity.

3. Method: A quasi-experimental questionnaire-survey based on a mixed-method design of Clift and Hancox (2001) was employed. Firstly, qualitative and quantitative data were obtained by a music and health questionnaire, the WHOQOL-Bref and a music school evaluation scale with answers from adult layman instrumental students (N=71) from 12 music schools. Secondly, in pursuance of a qualitative content analysis a reliable 26 item questionnaire was developed to gather data about group comparisons and sub-scale correlations.

4. Results: Most participants mentioned benefits of emotional well-being (94%). Thereby they focus especially on improvements to overcome daily routine challenges. Relaxation, distraction, and balance are frequently mentioned categories, all refer to a succesful use of an alternative musical activity help to manage exhausting everyday life tasks. Social well-being was rated higher for instrumental students receiving group lessons than for those taking private lessons ($p<.05$). However, no significant correlations between sub-scales of quality of life or music school quality areas and scores of well-being domains were found.

5. Conclusions: Concerning emotional and social well-being, the present study supports the findings of Clift and Hancox (2001) research on choir singing. The data suggest that making music together enhances adult instrumental music students' experience and feeling of social well-being, but this should be proved in further research.

[2-04] THE ADAPTIVE FUNCTIONS OF MUSIC LISTENING: AN EXPLORATORY FACTOR ANALYSIS

Jenny Groarke[1], *Michael Hogan*[1]
[1]*School of Psychology, National University of Ireland, Galway, Ireland*

Poster

1. Background: This paper details the development of the Adaptive Functions of Music Listening Scale. The current study draws upon psychological theory (e.g. affect regulation, well-being) to explain music listening motivations. Current measures are limited by failing to consider both hedonic and eudaimonic motivations in music listening, which may have differential associations with well-being.

2. Aims: The current study aims to develop a scale to measure functions of music listening that is suitable for analysing theoretically-driven research questions in relation to the affective, social, and cognitive consequences of different music listening strategies. Analysis of music listening motivations and the impact of listening strategies are poorly developed as a consequence of limited theory and scale development in this particular area.

3. Method: Item generation was on the basis of an exhaustive literature review, and four qualitative focus group sessions. Thematic analysis revealed 6 broad dimensions of music listening (affective, social, cognitive, everyday listening, eudaimonic motivations, and identity functions). Items (N = 240) were rated by three content experts in terms of their coherence and relevancy. Only items rated highly by all experts were retained. 150 items were administered to a sample of 500 undergraduate students, for exploratory factor analysis (EFA) and further scale refinement.

4. Results: A complex, higher-order factor structure is hypothesised, with each dimension of music listening being composed of 2-6 sub-factors. Factors will be retained based on results of Horn's parallel analysis. Items will be retained on the basis of item-total correlations greater than .30 and factor loadings in excess of .50. Items cross-loading at .32 or higher will be removed. Data analysis is ongoing, and results will be presented in August.

5. Conclusions: At present there are many applications of music. Understanding the functions of music listening may help explain why music is of value in certain contexts. Music listening motivations vary within and between individuals, and may have important implications for outcomes in research and clinical practice.

[2-05] ANALYSIS OF VERBAL DESCRIPTORS FOR TIMBRE: BASED ON CONCEPTUAL METAPHOR THEORY

Yo-Jung Han[1]
[1]*The Pennsylvania State University, USA*
Poster

1. Background: Language about music can serve as an important indicator of our perception and cognition of music. Our verbal descriptions about musical parameters frequently rely on metaphor. Metaphor entrenched in everyday life is used without conscious effort, but it governs our thoughts (Kövecses, 2010). In this sense, understanding metaphor is a way of understanding our thoughts and experiences. According to conceptual metaphor theory, when A is understood as B, the target A is conceptualized by the source B. It has been found that the source domain of musical parameters such as pitch, duration, and loudness is spatial dimension (Moravec & Stepanek, 2005). However, less attention has been given to the examination of timbre.

2. Aims: This study aims to understand how we conceptualize timbre by investigating the source domain of timbre descriptors based on conceptual metaphor theory.

3. Main contribution: As a result of the semantic analysis of sixty-nine descriptors presented in three previous studies on timbre, I found vision, touch, sound, taste, and person were used as the source domains of timbre. The dominant source domains were vision and touch. The subcategories of vision were brightness, clarity, density, shape and spatial dimension. Touch had surface and solidity of materials, temperature, and humidity as the subcategories. In previous research, frequently used domains were brightness, clarity, spatial dimension, and density in vision and surface and solidity in touch. Timbre was conceptualized based on various source domains, as opposed to pitch, duration, and dynamics.

4. Implications: Understanding the source domains of timbre descriptors is a way of understanding how we conceptualize timbre through a more concrete bodily experience. The study yields a new question: why certain subdomains such as brightness, surface, and clarity were frequently used as compared to other domains. Hence, it seems necessary to investigate the relationship between the physical attributes of timbre and certain source domains.

[2-06] THE RELATIONSHIP BETWEEN MUSIC AND LANGUAGE IN TERMS OF DERRIDA'S CONCEPT OF WRITING

Hye Kyung Park[1]
[1]*University At Buffalo, USA*
Poster

ABSTRACT

I examine the limitations in trying to seek meaning by continuing to engage the traditional relationship between language and music, and propose a new relationship between them in terms of Derrida's concept of writing. I begin with Derrida's deconstruction of the binary opposition, speech/writing, in language. Derrida states that writing comprehends language, whereas for Rousseau, it is only the simple "supplement to the spoken word." Derrida expanded the concept of writing to include any type of inscription. Based on his new construction of writing/speech, language represents the creative nature of the human mind, found in writing. Thus, it contains the connotative function of language, and using Derrida's concept of writing, I propose music as a language.

1. INTRODUCTION

While seeking to explain meaning in music, musicological and philosophical discussions usually begin with the traditional premise of the binary in which language is the dominating element, and music its subordinate. However, Iegor Reznikoff states that "the level of sound is much more primitive in our consciousness than the level of speech." He believes that sound is related to the very first levels of consciousness, which already appear in the period before birth. The level of language, on the other hand, is specialized, which a child only begins to acquire after birth attaining a full grammar at around the age of three. Despite the existence of this primal characteristic of music in the brain, it is still considered to be subordinate to language. Therefore, why is it that language is thought to be superior to music? In this paper, I examine the limitations in trying to seek meaning by continuing to engage the traditional relationship between language and music, and propose a new relationship between them in terms of Jacques Derrida's concept of writing.

2. JACQUES DERRIDA'S CONCEPT OF WRITING

My deconstructive approach to music is based on Derrida's examination of the relationship between speech and writing. Deconstruction, as a mode of interpreting a text, led Derrida to consider the relationship of speech and writing in Rousseau's Essay on the Origin of Languages. Rousseau thought writing to be language in the absence of a speaker. Raymond Monelle summarizes the traditional thought about the relationship between the two as the following: "speech is a direct representation of the thought of the speaker; its meaning is manifestly present. But writing is a representation of a representation. It is a supplement to speech, sinister and elusive, an unreliable counterfeit."30 However, Derrida states that writing comprehends language, since he analyzes the hierarchy of the binary as undecidable by way of a detailed investigation of the relations between the two apparently opposed terms of the structure. In the case of speech/writing, the relation between the two is more complicated than is implied by the terms "representation" and "dependence." He asserts that "Either writing was never a simple 'supplement,' or it is urgently necessary to construct a new logic of the 'supplement.'"[31]

Interestingly, Derrida expands this concept of writing to other areas, such as music, art, and dance. He states:

30 Raymond Monelle, *Linguistics and Semiotics in Music* (Rondon: Routledge Press, 1992), 304.
31 Derrida, *Of Grammatology* (Baltimore: Johns Hopkins University Press,1976), 7.

> One says "language" for action, movement, thought, reflection, consciousness, unconsciousness, experience, affectivity, etc. Now we tend to say "writing" for all that and more: to designate not only the physical gestures of literal pictographic or ideographic inscription, but also the totality of what makes it possible; and also, beyond the signifying face, the signified face itself. And thus we say "writing" for all that gives rise to an inscription in general, whether it is literal or not and even if what it distributes in space is alien to the order of the voice: cinematography, choreography, of course, but also pictorial, musical, sculptural "writing."32

Broadly, any type of inscription, according to Derrida, can be considered to be a form of writing. This implies that not only does the system of notation secondarily connect with these activities, but also their essence and content. In other words, painting, sculpture, cinematography, choreography, writing, and musical composition belong to the big category, WRITING. That is, writing is no longer a subcategory of language; writing is no longer restricted to language. In this respect, I will expand his theory to encompass music.

3. THE RELATIONSHIP BETWEEN MUSIC AND LANGUAGE

3.1. Music as a Universal Language

Before my discussion of music in terms of writing, let us first explore the relationship between music and language. First, music is thought to be a universal language in the form of representation. Rousseau believed that both music and speech are languages in that they are expressive. According to Hermerén, "all languages allow speakers (1) to make propositions, in other words, to refer to specific entities or concepts (such as "cousin" or "justice") and to predicate things about them; (2) to express wishes and ask questions; and (3) to make metalinguistic statements, such as "In my language we do not have a word that means exactly the same thing as 'justice' in your language."33 However, music does not bear this type of meaning. Despite the fact that they are different in aspect, there is also a common aspect. Language is not only denotative, but also connotative. In other words, if I call my friend a genius, it could mean that she is really one or I may be speaking sarcastically. Meaning does not exist only on the symbolic level of language; it depends on the pitch and contour of a voice. Like language, music represents meaning, but primarily connotatively rather than denotatively. Thus, meaning in music differs according to the listener. Therefore, when we think about poetic language, both music and language are the same in terms of connotation. Based on this aspect of representation, David Lidov states, "music is a universal language of the emotion."34 Regarding universality, those who travel abroad may have this kind of experience; even though s/he cannot understand a foreign language on TV, s/he can sense the meaning of the music, composed by a foreign composer. Music, then, crosses cultural boundaries more easily than does language. As another example, Korean popular music has currently spread on-line to Asia, Europe, and North and South America. Although the people of these countries do not understand the lyrics of Korean pop, they can become absorbed in the music, eventually intuiting the meaning of the lyrics. In other words, this music speaks to the world, and the world begins to comprehend. Lidov presents a similar example as follows:

> . . . [w]e still see much evidence that people learn new musics more rapidly than new languages, at least to a level of competence that permits engaged and contrasted listening experiences. When Debussy heard a Javanese gamelan at the Paris Exhibition of 1898, he got something he could take home. I wouldn't suggest that he had acquired expertise, but he and his compatriots evidently felt a significance and richness in gamelan music – without translation – that was not available to them in untranslated Indonesian poetry.35

Thus, music can be considered to be a language, in particular, a universal language.

3.2. The Similarity of Function between Music and Language in the Origin

The second view of the relationship of music and language seeks their similarity of function. In his Essay on the Origin of Languages, Rousseau states that "verse, singing, and speech have a common origin."36 In the chapter titled "The Origin of Music and Its Relations," Rousseau describes the origin of music as follows:

> With the first voices came the first articulations or sounds formed according to the respective passions that dictated them. Anger produces menacing cries articulated by the tongue and the palate. But the voice of tenderness is softer: its medium is the glottis. And such an utterance becomes a sound. It may occur with ordinary or unusual tones, it may be more or less sharply accented, according to the feeling to which it is joined. Thus rhythm and sounds are born with syllables: all voices speak under the influence of passion, which adorn them with all their éclat. Thus verse, singing and speech have a common origin. Around the fountains of which I spoke, the first discourses were the songs. The periodic recurrences and measures of rhythm, the melodious modulations of accent, gave birth to poetry and music along with language.37

Rousseau speculates, then, that the first tales, the first speeches, and the first laws were in the form of verse. Poetry was devised before prose, because "feelings speak before reason." Music did not exist at first, but melody did. Melody was in the form of the varied sounds of speech. "Accents constituted singing, and quantity constituted measure, and one spoke as much by natural sounds and rhythm as by articulations and words." Thus, verse, singing, and speech in the earliest societies were bound together. The language of the Pirahã, a small tribe from the Brazilian Amazon supports this assertion. Daniel and Keren Everett researched this tribe for thirty years, and discovered that the members of this tribe speak a language which possesses one of the smallest phoneme inventories of any known language. Pirahã has no concept of number, and no fixed term for colors other than the term of light and dark, no perfect tense, no deep memory, no tradition of art or drawing, and no words for "all," "each," "every," "most," or "few."38 Moreover, Everett's most explosive claim was that their language displays no evidence of recursion

32 Derrida, *Of Grammatology*, 9.
33 Aniruddh D. Patel, *Music, Language, and the Brain* (Oxford: Oxford University Press, 2008), 301; G. Hermerén, "Representation, Truth, and the Languages of the Arts," in *Essays on the Philosophy of Music, Acta Philosophica Fennica* (Vol. 43), eds. V. Rantala, L. Rowell, and E. Tarasti (Helsinki: The Philosophy Society of Finland), 179-209.
34 David Lidov, *Is Language a Music?: Writings on Musical Form and Signification* (Bloomington, IN: Indiana University Press, 2005), 1.
35 David Lidov, *Is Language a Music?: Writings on Musical Form and Signification*, 2.
36 Jean-Jacques Rousseau and Johann Gottfried Herder, *On the Origin of Language* (Chicago: The University of Chicago Press, 1986), 50.
37 Jean-Jacques Rousseau and Johann Gottfried Herder, *On the Origin of Language*, 50.
38 John Colapinto, "The Interpreter: Has a remote Amazonian tribe upended our understanding of language?" http://www.newyorker.com/reporting/2007/04/16/070416fa_fact_colapinto#ixzz1hwLPzemI (accessed 12/27/2011).

or embedding one clause within another.39 Because they do not have the concept of numbering and counting, their language reveals a length limit. In other words, Pirahã is the one of the simplest language systems. However, "it possesses such a complex array of tones, stresses, and syllable lengths that its speakers can dispense with their vowels and consonants altogether and sing, hum, or whistle conversations."40 Thus, music is overtly part of their language. Their simple sentences are like verses, and communication through song reveals the relationship of music and language. Pirahã, although it is not a primitive language, could exemplify the earliest societies providing an affirmative answer to Lidov's question, "is language a music?"

3.3. The Connection in terms of Cognitive and Neutral Procession beneath the Difference between Music and Language

Recently, Aniruddh D. Patel compared music and language in his Music, Language, and the Brain. In his comparison of sound elements between the two domains, he states that, on the surface, they are dramatically different. "Music uses pitch in ways that speech does not, and speech organizes timbre to a degree seldom seen in music."41 But beneath this difference, there lies a deep connection in terms of cognitive and neutral processing. "In both domains the mind interacts with one particular aspect of sound (pitch in music and timbre in speech) to create a perceptually discretized system."42 He concludes, "Although music and speech differ in the primary acoustic feature used for sound category formation, it appears that the mechanisms that create and maintain learned sound categories in the two domains may have a substantial degree of overlap."43 I believe that Schoenberg's *Sprestimme*, in which the melody is spoken at approximate pitches rather than sung on the exact pitches, exploits this relation between spoken and musical sound systems.

Furthermore, Patel compares the rhythmic aspect of music and language. Rhythm in music is a very important element, but in speech it has gone unnoticed. He states that "[s]peech and music involve the systematic temporal, accentual, and phrasal patterning of sound." Their rhythms also show important similarities and differences as follows:

> One similarity is grouping structure: In both domains, elements (such as tones and words) are grouped into higher level units such as phrases. A key difference is temporal periodicity, which is widespread in musical rhythm but lacking in speech rhythm. Ironically, the idea that speech has periodic temporal structure drove much of the early research into speech rhythm, and was the basis for a rhythmic typology of languages which persists today (stress-timed vs. syllable-timed languages). It is quite evident, however, that the notion of isochrony in speech is not empirically supported. Fortunately, much recent empirical research on speech rhythm has abandoned the notion of isochrony, and is moving toward a richer notion of speech rhythm based on how languages differ in the temporal patterning of vowels, consonants, and syllables. A key idea that motivates this research is that linguistic rhythm is the product of a variety of interacting phonological phenomena, and not an organizing principle, unlike the case of music.44

4. MUSIC AS A UNIVERSAL LANGUAGE VS. MUSIC POSSIBLY AS LANGUAGE

As described above, music scholars often have compared music and language. Their relationship can largely be divided into two views: the traditional thought about the relationship between the two, music as a universal language, and the recent view by scholars such as Lidov, music possibly as language. In the first view, language is considered to be more fundamental than music. Thus, while music theorists discuss various aspects of music as an allegory of language, linguists rarely study language in terms of music. In the same vein, in chapter 6 of Music, Language, and the Brain, Patel restates the traditional thought about music and language when he begins by stating that "language" refers to the ordinary language of everyday communication, not poetry, philosophy, or other specialized forms of discourse. "Music" refers to instrumental music – music without words – unless otherwise stated." Patel mainly focuses on the denotative aspect of language, and, unfortunately, this is what he compares to music. Patel, then, states that music does not bear denotative meaning, but "[m]eaning exists when perception of an object/event brings something to mind other than the object/event itself."45 Here, he is in agreement with this Jean-Jecques Nattiez's position of semiotic meaning. It is contrary to Kivy's definition of meaning that "it should be reserved for the linguistic sense of reference and predication;" in this view, music does not have meaning. Thus, Patel's argument seems progressive in viewing the relationship between the two, especially in his comparison of pitch, rhythm, melody and syntax in music and language. However, Patel's research reverts to the conventional point of view of language as ordinary speech. In this traditional view, music can only be the supplement of language, possessing a few characteristics found in language in terms of simply and vaguely giving connotative and emotional meaning, as seen in Kivy's definition. This conventional view is based on the idea that language is understood as speech (writing, a mere supplement). Thus, speech is said to be superior to writing. However, as long as this construction is maintained, language (as speech) appears to be superior to music, and it becomes difficult to know not only the substance of language, but also that of music. Thus, the power of music becomes diluted.

Meanwhile, Lidov's recent study about the relationship between language and music raises the status of music to an equal position with language. His question, "If music is a universal language, is language a universal music?" suggests the reversal of the conventional recognition about music and language; music is a kind of language, and language is a kind of music. Although the musical aspect of speech is evident, on the simplest level, in its prosody, the music of language has not yet been researched. Thus, he suggests the Derridean inversion of these binaries, music and language. Then, he concludes that "some of the things we do not understand about music and some of the things we do not understand about language may well be the same."46 Even though, like much conventional linguistic thought, he still thinks of language in terms of speech, his assertion is valuable for rethinking the relationship between these two domains.

5. MUSIC AS A LANGUAGE

However, this traditionally fixed idea of language as speech restricts the meaning of language, and then, in turn, the meaning of music. Speech is limited only to communication, which is merely a denotative function of language. Like music, language represents the creative nature

39 Daniel Everett, "Cultural constraints on Grammar and Cognition in Pirahã: Another Look at the Design Features of Human Language," *Current Anthropology* 46 (2005): 628.
40 John Colapinto, "The Interpreter: Has a remote Amazonian tribe upended our understanding of language?"
41 Aniruddh D. Patel, *Music, Language, and the Brain*, 86.
42 Aniruddh D. Patel, Music, Language, and the Brain, 86.
43 Aniruddh D. Patel, Music, Language, and the Brain, 86.
44 Aniruddh D. Patel, *Music, Language, and the Brain*, 176-7.
45 Aniruddh D. Patel, *Music, Language, and the Brain*, 304.
46 Lidov, *Is Language a Music?*, 14.

of the human mind found in writing. As Derrida asserts in Of Grammatology, language can be best understood as writing. Therefore, in terms of writing, I propose that music is a language. As seen in the early societies mentioned by Rousseau and the example of Pirahã, the assertion that music and language may be the same can be seen in the strong connection between poetry and music. Of course, musical meaning cannot represent a single concept because it is not denotative. And yet, when we expand our concept of language to poetry and philosophy, we begin to understand how music and language may be the linked. Poetry, as connotation, is the representative form of language; it is writing. Like music, it is primarily governed by idiosyncratic form and convention to manifest the differential meaning of words, layered meaning. Depending on the context and reader, a work carries different meanings. There is no single meaning. While Patel, in his traditional analysis, excludes poetry, philosophy, and other specialized forms of discourse in the comparison of language and music, I believe, in contrast, that music actually falls into these specialized forms. Like poetry, music also does not restrict the interpretation of its meaning. Furthermore, meaning in music does not exist only in terms of emotion. For example, the listener can find meaning in the play of the subject and answer in Bach's fugues. Thus, like poetry, it is connotative and open to interpretation. Therefore, music as a form of writing greatly expands our traditional thought of language as predominately denotative to language as connotative, and, therefore, includes music as a language. This position is in full agreement with Derrida's expanded concept of writing — any type of inscription can be considered as writing, such as painting, choreography, sculpture, musical composition. Therefore, music and language appear to be the same in terms of this concept of writing or composition.

6. CONCLUSION

Jacques Derrida's deconstruction of speech and writing extends itself to that of language and music. When we invert the conventional binary of speech/writing to writing/speech, the expanded concept of writing, any type of inscription, allows for the reconsideration of the traditional relationship of language/music. Thus, music is a language, since language, in terms of writing, is connotative as is music. This logic will not dilute the power of music and will help to better understand its substance.

7. REFERENCES

Colapinto, John. 2007. "Interpreter: Has a remote Amazonian tribe our understanding of language?" http://www.newyorker.com/reporting/2007/04/16/

Derrida, Jacques. 1976. Of Grammatology. *Baltimore: Johns Hopkins University Press.*

Everett, Daniel. 2005. "Cultural constraints on Grammar and Cognition in Pirahã: Another Look at the Design Features of Human Language," Current Anthropology *46: 621-646.*

Kivy, Peter. 1993. The Fine Art of Repetition: Essays in the Philosophy of Music. *Cambridge and New York: Cambridge University Press.*

Lidov, David. 2005. Is Language a Music?: Writings on musical form and Signification. *Bloomington and Indianapolis: Indiana University Press.*

Monelle, Raymond. 1992. Linguistics and Semotics in Music. *London: Routeledge. Patel, Aniruddh. 2008.* Music, Language, and the Brain. *Oxford and New York: Oxford University Press.*

Rousseau, Jean-Jacques and Herder, Johann Gottfried. 1986. On the Origin of Language. *Chicago: The University of Chicago Press.*

[2-07] MUSIC TRAINING AND LINGUISTIC STRUCTURE PROCESSING IN CHILDREN AND ADULTS

Swathi Swaminathan[1], Glenn Schellenberg[1]
[1]University of Toronto, Canada

Poster

Music, like language, contains sounds arranged in rule-governed structures. Training in music, therefore, involves implicitly learning about musical structures and producing them. It is not known whether the effects of music training also extend to the processing of linguistic structure. In the present work, we examine the association between training in music and processing linguistic structure in children and adults. In one experiment, musically trained and untrained primary-school children were compared on their ability to process grammatical structure. The children also completed tests of verbal working memory and IQ. In a second experiment, we tested trained and untrained undergraduates on linguistic structure processing, verbal working memory, and IQ. In both experiments, the musically trained group outperformed their untrained counterparts on the verbal working memory test. No group differences were found for linguistic structure processing. This result suggests that the linguistic benefits of music training, if any, are likely limited to basic skills involving auditory and general cognitive abilities rather than higher level language skills such as processing grammar.

[2-08] BRANCH-DEPENDENT TIME- AND TIMBRE-RELATED FEATURES OF AUDIO LOGOS

Isabella Czedik-Eysenberg[1], Christoph Anzenbacher[1], Christoph Reuter[1], Michael Oehler[2]
[1]University of Vienna, Musicological Department (systematic Musicology), Austria, [2]Macromedia University for Media and Communication Cologne, Germany

Poster

In the recent years audio logos became also more and more the focus of music psychologic attention. In a number of studies audio logos have already been examined concerning their melodic comprehensibility, their impact in cross modal perception, their noticeability in difficult auditory environments etc. pp. (North et al. 2004; Spence/Driver 2004, Allan 2007; Bronner/Hirt 2007; Anzenbacher 2012; Anzenbacher/Reuter/Oehler 2013; Langeveld et al. 2013). However the question of typical acoustical features of audio logos remained still unanswered. Therefore in this current study the following questions are in the centre of attention: what are typical acoustical features of audio logos? Are there certain acoustic attributes, which are especially peculiar to the audio logos of certain industrial branches? In a preliminary study with 127 audio logos from different industrial sectors their auditory features like spectral centroid, inharmonicity, roughness, dynamic range, length, mean centroid pitch ratio, key, instrumentation/voicing etc. have been investigated via Matlab/MIRtoolbox (Lartillot/Toiviainen 2007). In a subsequent step the study was expanded to 364 audio logos and additional parameters (first attack time, unpleasantness, melodic contour, presence of human voice/specific instruments) were taken into account. The audio logos were compared/correlated with each other in terms of their individual industrial branch by means of a ANOVA and Tukey post-hoc analysis. Within the branch "automotive/automotive services" audio logos tend to have higher roughness as well as inharmonicity. Audio logos of the branches "laundry/home care/beauty", similar to "retail", seem to be relatively short while having a higher spectral centroid as well as spectral fluctuations and roughness. In contrast to that, low

roughness, spectral fluctuations and a low spectral centroid can be observed for audio logos of educational institutions and organisations, which furthermore tend to have audio signatures of noticeably longer duration. In respect to the melodic contour of an audio logo mostly wavelike and ascending structures can be found. Besides to spoken and sung slogans, whispering is also common in audio logos, especially in the areas of "retail" and "food". It can be shown that audio logos have typical auditory and musical features with some notable characteristics depending on different industrial branches. The approach of semi content based music information retrieval offers opportunities for the classification of short audio sequences such as audio logos. For an automated semantic audio analysis, further research on a larger sample would be needed, in order to fully give consideration to the content and music psychological impact of audio logos.

[2-09] PHYSIOLOGICAL AND ACOUSTICAL CORRELATES OF UNPLEASANT SOUNDS

Christoph Reuter[1], Michael Oehler[2], Jörg Mühlhans[1]
[1]University of Vienna, Musicological Department, Austria, [2]Macromedia University for Media and Communication, Germany
Poster

1. Background: There are many sounds that people perceive as unpleasant that may cause physiological reactions. At least since 1959 the “pleasantness” of sound is discussed from different angles (Boyd 1959; Ely 1975; Halpern et al. 1985; Neumann & Waters 2006; Cox 2008; Grewe et al. 2010). Frequently scratching fingernails on a chalkboard or squeaking chalk top the list of unpleasant sounds.

2. Aims: The aims of our study were (a) to find specific acoustic features of sound responsible for perceived unpleasantness, (b) to examine the correlation between perception and physical reaction and (c) to find potential influences of knowledge about the origin of sounds or the method of exposure on perception and physical reactions.

3. Methods: All subjects (N=96) hat to rate the original sounds of scratching fingernails, styrofoam, a squeaking instrument and vomiting as well as 6 variations of them: high-, low-, bandpass and -stop or extracted tonal/noise components. During listening, measurements of galvanic skin response and muscle tone were conducted with a Nexus-10 biotrace device.

4. Results: The most significant result was the influence of pitch. Sounds with deleted pitch information were rated significantly more pleasant ($p<0.001$ for instrument and fingernail) whereas deleted noise parts did not influence the ratings. For the vomiting sound the opposite was true in both cases. Furthermore the frequency band 2-4 kHz also seems to be important for the perceived unpleasantness. Deleting those frequencies led to significantly better ratings ($p<0.001$ for fingernail, styrofoam and instrument) while again for vomiting no significance was found. Correlations between galvanic skin response and subjective rating were not significant, still the curves matched to a large extent except for the vomiting sound. No significance could be found for the method of exposure and the origin of sounds.

5. Conclusions: The relevant acoustic characteristics of unpleasant sounds can be found in the pitch information and the frequency range between 2000 Hz and 4000 Hz, where the ear is most sensitive. Such (unpleasant) sounds seem to evoke a physical reaction in the subjects. For sounds like vomiting rather the association (disgust) leads to negative ratings than the acoustic features.

[2-10] ROLE AND EFFECTS OF PROCEDURAL ORIENTATION ON SKILL ACQUISITION IN DELIBERATE MUSIC PRACTICE

Frank Liebscher[1]
[1]Hochschule Für Musik & Theater "felix Mendelssohn Bartholdy" Leipzig, Germany
Poster

1. Background: In everyday life musicians of all genres and levels facing multiple challenges practicing their instruments. As showed by research of related sciences, attaining higher levels of performance strongly depends on the degree of automatization of a musician’s sensori-motor and cognitive skills. Accordingly, most practice methods suggest start practicing in lower tempi – usually from the beginning of the musical piece, section or phrase. Nevertheless, the design of anterograde practice (AP) seems to bear some problematic implications, which may impair the efficiency of the musical learning process. Consequently, Retro Sequential Practice (RSP) has drawn some attention as a music practice method, that is trying to approach those problems alternatively by focusing on the primary automation of the terminal sequences of a musical object and its stepwise (sequential) backward (retro) oriented expansion.

2. Aims: Accordingly, the talk will outline the role and effects of anterograde/retrograde orientation on skill acquisition in music practice, explain the effects regarding to neuroscientific findings and introduce options and possible benefits of applying RSP for optimizing deliberate music practice.

3. Methods: The study is conducted in form of an empiric field study, which is still in progress and includes students of differing ages, levels of expertise and musical goals. All participants were tested on RSP and AP. The sessions were videotaped and supplemented by oral interviews. First, the analysis of the footage helped to indicate the main problems in AP. Furthermore, to point out the distinct differences of AP and RSP, their key features were portrayed, analyzed and discussed. Finally, in order to explain the possible advantages of RSP, the empiric findings were interpreted and discussed regarding to neuroscientific research results.

4. Results: The findings and their neuroscientific interpretation suggest, that RSP supports the musical learning process, particularly regarding to accomplishing of complexity and structuring of learning content, specifying of perception and sharpening of imagery, enhancing of retention and accelerating of skill development, strengthening of self-confidence, intrinsic motivation and improving stage presence.

5. Conclusion: Overall, RSP optimizes the musical learning effect by conclusively aligning deliberate practice according to neurophysiological and neuropsychological facts, conditions and phenomena.

[2-11] MUSICAL, LANGUAGE, AND SOCIAL DEVELOPMENT OF ONE-YEAR-OLDS PARTICIPATING IN MUSIC CLASSES

Sachiyo Kajikawa[1], Hideo Moriuchi[2]
[1]Tamagawa University, Japan, [2]Yamaha Music Foundation, Japan
Poster

ABSTRACT

This study examined the influence of music lessons on the development of one-year-old children. The musical and general development of children who had music lessons was compared with those who did not. Participants were one-year-old children and their parents and were assigned to either a music group or to a control group who received no lessons. The music classes for the music group were held every two weeks for the duration of one year. Parents completed a questionnaire about parent-child communication at home, and the musical and general development of their child three times within an interval of six months. During the year of the study, both the frequency and duration with which parents sang and listened to music with their children were higher in the music group than in the control group. The music group also had higher scores than the control group for movement and listening at all three assessments. At the first and second assessments, there were no significant differences between the two groups in singing, but the music group had higher scores in this domain than the control group at the third assessment. On the assessment of general development, the two groups did not differ in productive language at first. However, the music group had higher scores than the control group at the second and third assessments. Manipulation behaviors, language concepts, and social skills with other children and adults also showed a similar pattern. The experience of participating in music lessons, as well as musical communication at home, may promote the development of one-year-old children, especially with regard to singing, language, and social behavior.

1. INTRODUCTION

Before attending kindergarten or nursery school, infants and children experience music mainly in their home. However opportunities for young infants to experience music outside their home have been increasing in recent years.

It is well known that musical experiences influence infant and child development in various ways. For example, children who received music lessons for more than three years had a larger lexicon than those who did not (Piro & Ortiz, 2009). Another study indicated that children who participated in group music activities at school engaged in more conversation with their parents, and had stronger self-esteem and higher motivation for activity than those who did not (Hallam, 2010). One study focusing on IQ indicated that keyboard and voice lessons promoted 6-year-old participants' cognitive development, whereas drama lessons had an effect on their social skills (Schellenberg, 2004). Based on this result, the authors suggested that enjoyable small-group lessons would promote children's general abilities, an effect not limited to music education.

It has also been shown that the type of music lessons an infant participates in has an effect on the development of music cognition, social skills, and communication skills (Anvari, Trainor, Woodside, & Levy, 2002). Musical activities may promote infant-parent communication at home, which influences infant development more broadly. However, few studies have been conducted on children younger than three years of age.

This study examined the influence of music lessons on the development of one-year-old children. The children participated in group music lessons with their parents for one year. We examined their development three times: at the beginning, the middle, and the end of the lesson period. Musical development and general development were compared between children who had music lessons and those who did not.

2. METHODS

2.1. Participants

Participants were one-year-old children (N = 332, Mean Age = 1 year and 6 months) from five major cities in Japan: Tokyo, Nagoya, Sapporo, Osaka, and Fukuoka. Participants were assigned to either a music group (n = 87; 45 girls, 42 boys), who participated in music classes regularly, or to a control group (n = 245; 144 girls, 131 boys), who did not experience music lessons. All the children were healthy first-borns and lived with their parents.

2.2. Procedure

Music classes were held every two weeks for groups of eight children over one year. All the teachers had received professional music education and had the technical skills necessary to teach music to small children.

In 45-minute classes, children listened to piano playing by a teacher, sang songs together, played using non-musical tools such as cloth while hearing music, listened to picture books with background music, and moved with music. Each activity was conducted for about five minutes and parents participated in all of the activities with their child.

During the research period of one year from April 2012 to March 2013, parents completed a questionnaire three times within an interval of six months: the 1st assessment around the start of the research period, the 2nd around the middle, and the 3rd around the end.

The parents were first asked questions regarding parent-child communication at home, such as the frequency, time length, and content of singing and listening to music, as well as playing with toys and reading picture books.

Then, they assessed their child's development using a checklist of behaviors extracted from the Kinder Infant Development Scale (KIDS; Miyake, 1991). The list included motor behaviors, manipulation, language concepts, receptive and productive language, social skills with adults and other children, and daily activities. There were seven items for each domain, that is, 49 in total.

Parents also completed a checklist regarding musical behaviors, such as listening, moving (dancing), and singing. The items were based on child musical development at the age of 0 – 3 years (Tanimura, 2010). The number of items was three, two and five respectively. The total number of items checked by the parent in each domain was analyzed as a developmental score.

3. RESULTS

Both the frequency and duration with which parents sang with their children were higher in the music group than in the control group during the year of the study (Fig.1). The frequency and duration did not change significantly throughout the period for either group. The frequency and duration with which children listened to music at home were also higher in the music group than the control group (Fig.2). Children in the music group had more music experience at home than those in the control group and this difference was maintained until children were at least two years of age.

In all three assessments of musical development, the music group had higher scores than the control group for listening (ANOVA, $F(1,330) = 308.2$, $p < .001$, Fig.3) and movement ($F(1, 330) = 467.6$, $p < .001$). For singing, the main effect of time ($F(2, 660) = 109.1$, $p < .001$) and the interaction between group and time ($F(2, 660) = 76.8$, $p < .001$) were also significant. In the first and second assessments for singing, there was no significant difference between the two groups. However, in the third assessment, the music group had higher scores than the control group ($p < .001$).

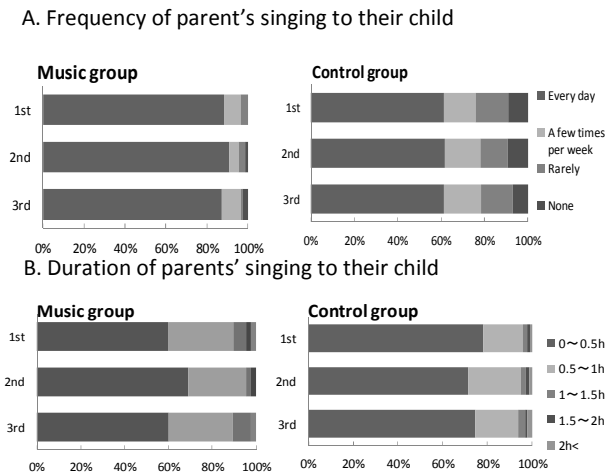


Fig.1 Frequency and duration with which parents sang to their child in the music and the control groups.

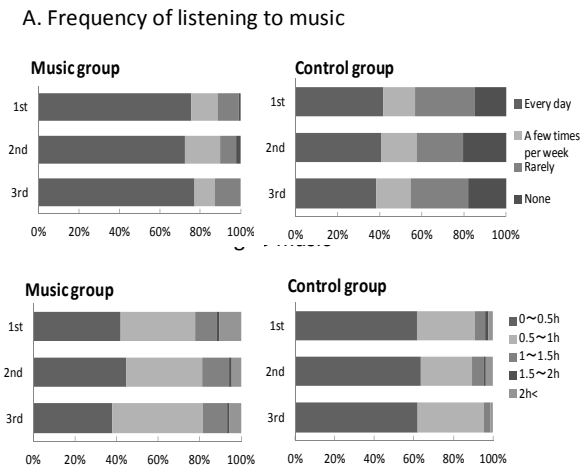


Fig.2 Frequency and duration with which parents listened to music with their child in the music and the control groups.

Among the assessments of general development, only motor behaviors showed no significant interaction between group and time ($F(2,660) = 1.7$, *ns*) but all other domains did (manipulation: $F(2,660) = 11.7$, $p < .001$; language concepts: $F(2,660) = 24.4$, $p < .001$; receptive language: $F(2,660) = 12.2$, $p < .001$; productive language: $F(2,660) = 7.0$, $p < .01$; social skills with children: $F(2,660) = 13.2$, $p < .001$; social skill with adults: $F(2,660) = 24.5$, $p < .001$). In these domains, there were no differences between the two groups at the first assessment, although the music group acquired a higher mean score than the control group at the second and third assessments (Table 1). Only on productive language was there no main effect of group ($F(1,330) = 3.5$, *ns*).

Table 1 Mean (*SD*) of general development scores

Domain		Music group	Control group
Motor	**1st**	**0.86 (1.08)**	**0.73 (1.04)**
	2nd	**1.73 (1.59)**	**1.25 (1.40)**
	3rd	**2.57 (1.82)**	**2.09 (1.69)**
Manipulation	**1st**	**1.29 (1.15)**	**1.36 (1.00)**
	2nd	**2.53 (1.74)**	**1.84 (1.25)**
	3rd	**3.52 (1.63)**	**2.68 (1.53)**
Concepts	**1st**	**1.37 (1.70)**	**1.45 (1.13)**
	2nd	**2.89 (2.37)**	**1.91 (1.44)**
	3rd	**4.16 (1.84)**	**2.62 (1.71)**
Receptive language	**1st**	**3.03 (2.18)**	**2.17 (1.55)**
	2nd	**4.74 (2.04)**	**3.03 (1.58)**
	3rd	**5.72 (1.68)**	**3.72 (1.60)**
Productive language	**1st**	**0.86 (1.50)**	**1.15 (1.18)**
	2nd	**2.46 (2.19)**	**1.97 (1.97)**
	3rd	**3.89 (2.20)**	**3.16 (2.15)**
Social skills w/children	**1st**	**1.93 (1.85)**	**1.52 (1.10)**
	2nd	**3.28 (1.77)**	**2.07 (1.55)**
	3rd	**4.24 (1.92)**	**2.68 (1.69)**
Social skills w/ adults	**1st**	**1.44 (1.72)**	**1.68 (1.08)**
	2nd	**2.91 (1.92)**	**2.29 (1.40)**
	#rd	**4.18 (1.92)**	**2.93 (1.58)**

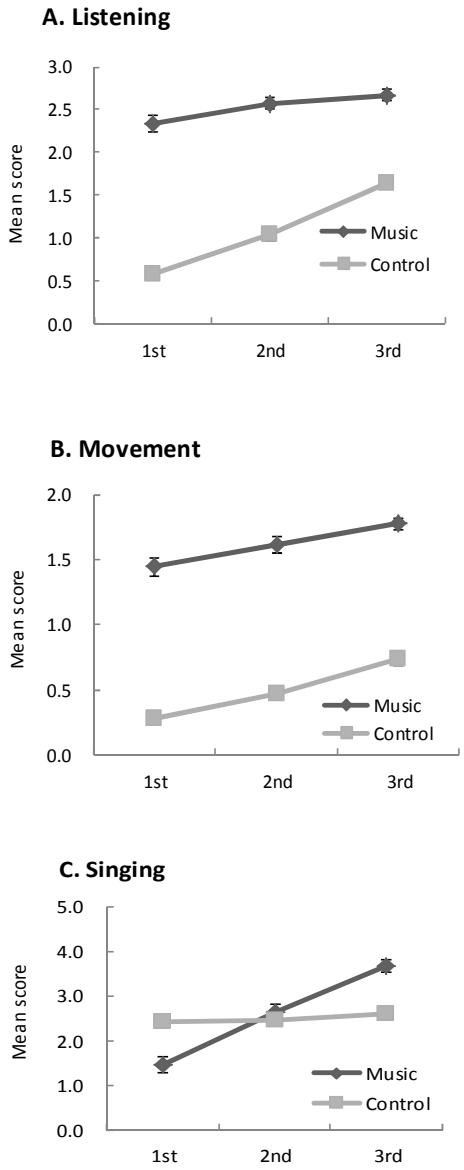


Fig.3 Musical development scores in the music and the control groups. Error bars indicate standard error.

4. DISCUSSION

Children who participated in music classes had more music experience with their parents at home before they started the lessons. In addition, these children showed more growth in listening and moving to music than those in the control group throughout the research period. There may be a relation between the quantity of music experience at home and children's response and sensitivity to music at this early age. Those parents whose children respond actively to music may believe they like music, increase their opportunities to play music at home, and decide to give them music education. Another interpretation is that music experience at home promotes children's response to music.

We found an effect of one year of music lessons on the development of singing behavior. Before the classes started, there was no difference between the two groups. Rather, the control group tended to score slightly higher on average than the music group. However, at the second and third assessments, children in the music group had higher scores than those in the control group. In every music class, children learned simple short songs repeatedly and sang in ensemble with the teacher, parents and other children. They were encouraged to sing songs using gestures that were matched to the music. These social and physical factors used to promote children's singing could be a relational factor in the group difference.

A group difference was also found in general development. The effect of music lessons was large, especially on productive language development, but all domains except motor behaviors were influenced by experience in music classes. This research examined the effects indirectly, but development of communication skills such as language and social skills with children and adults might have been promoted by music class participation. We cannot exclude the possibility that only regular experience in group activities outside the home may be a factor.

However, many previous studies indicated that music is a powerful tool for strengthening social bonds and social skills in both adults and children. Further research is necessary to investigate the direct relationship between music class activity and children's development in music and general domains.

In conclusion, experience in music lessons, as well as musical communication at home, may promote the development of one-year-old children, especially with regard to singing, language, and social behavior

5. REFERENCES

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. Z. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. Journal of Experimental Child Psychology, 83, *111-130.*

Hallam, S. (2010). The power of music: Its impact on the intellectual, social and personal development of children and young people. International Journal of Music Education, 28, *269-289.*

Miyake, K (1991) Kinder Infant Development Scale. *M. Omura, M. Takahashi, S. Yamauchi, and Y. Hashimoto (Eds.) Tokyo, Japan. Hattatsu Kagaku Kyoiku Center*.

Piro, J. M., & Ortiz, C. (2009) The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. Psychology of Music, 37, *325-347*.

Tanimura, H. (2010). Jiheishoji ni mirareru ongakukodo no henyo: Nyuyojiongakukodo no hattatsuprosesusukeru no sakusei o toshite (Transformation through music activities seen in a child with autism: Using a process scale of infant development through musical activities). Research on Early Childhood Care and Education, 48, *10-22.*

Schellenberg, E. G. (2004). Music lessons enhance IQ. Psychological Science, 15, *511-514.*

[2-12] MONTREAL BATTERY OF EVALUATION OF MUSICAL ABILITIES: EVALUATION OF CHINESE ADOLESCENTS' MUSICAL ABILITIES

Xue Tong[1]
[1]*Capital Normal University, China*

Poster

ABSTRACT

1. Background: Congenital Amusia is a lifelong impairment in musical ability, and deficit in distinguishing the fine-grained discrimination of pitch. The Montreal Battery of Evaluation of Amusia (the MBEA; Peretz, Champod, & Hyde, 2003) is currently the primary tool which is used for the identification of cases with congenital amusia across laboratories, but this battery has been validated for adults only. Detection of amusia in childhood is clinically important, so a tool for assessing musical abilities in children was developed: The Montreal Battery of Evaluation of Musical Abilities (MBEMA) and an abbreviated version in 2013 (Isabelle Peretz, Nathalie Gosselin & Yun Nan). The battery contains five tests: memory, scale, contour, interval, and rhythm. The battery was administered to 245 children (6–8 years of age) in Montreal and 91 children in Beijing. And its abbreviated version was administered to an additional 85 children in Montreal. The result shows that both versions of the MBEMA are sensitive to individual differences and to musical training.

2. Aims & Method: However, it has not been carried out with both Canadian and Chinese teenagers? The present study sought to verify the availability and stability of the abbreviated MBEMA in teenagers of China, as well as to assess the effect of musical training. 76 teenagers who were around 13-15 years old took the test.

3. Results & Conclusions: Based on the data analysis, the abbreviated MBEMA can identify effectively that teenagers who were judged as musical disorders (3 out of 76, approximately 4%). Then the three teenagers who might have musical disorders were retested by the full MBEMA. The result shows that they have been diagnosed with Congenital Amusia. However, there is no significant effect of musical training. Maybe the reason is that the battery is much easier for teenagers. So the abbreviated MBEMA can be used as a tool for diagnosing deficits in musical skills in children and adolescents of China; but it might not be an effective assessment of musical training for 13-15 years old Chinese adolescents, which may attribute to the mature abilities of tonal and rhymic perception at their age.

1. INTRODUCTION

Amusia is a musical disorder that appears mainly as a defect in processing pitch, but it also encompasses musical memory and recognition. Two main classifications of amusia exist: acquired amusia, which occurs as a result of brain damage, and congenital amusia, which results from a music processing anomaly at birth (Peretz et al., 2001). Congenital amusia is characterized by a deficit in musical pitch perception that cannot be explained by hearing loss, brain damage, intellectual deficiencies or lack of exposure. Amusia individuals have difficulty in recognizing familiar melodies and detecting out-of-key notes inserted in melodies (Peretz, 2008; Stewart, 2011). It is estimated that about 4% of the general population may have amusia, some of amusia individuals even years for music education.

In the past ten years, the study of congenital amusia was rapid developed. The research of pitch processing behaviour and neural basis, as well as the relation between amusia and language , amusia diagnose etc. have all had a great improvement.

The Montreal Battery of Evaluation of Amusia (MBEA) is a battery of tests that assesses musical abilities that has been developed and revised since 1987. The MBEA allows the diagnosis of different types of amusia by assessing musical abilities related to six components of musical processing presented in the neuropsychological model of musical cognitive processing, namely: Contour, Scale, Interval, Rhythm, Meter, and Musical Memory (Peretz et al., 2003). The MBEA has been used in studies of populations of individuals with brain injuries with different etiologies to assess various types of amusia and was shown to be useful for this purpose. Studies have used the MBEA to validate the battery. Satisfactory results were obtained from a psychometric perspective. Peretz et al. (2003) estimated that, although the data obtained for each test were asymmetric, the overall index followed a normal distribution and thus constituted a good index of perception and musical memory that can be used to distinguish between normal and deficient performance in the general population.

However, this battery has been validated for adults only (Isabelle Peretz,Nathalie Gosselin&Yun Nan, 2013). Detection of amusia in childhood is clinically important because of the greater malleability of developing brains (Huttenlocher, 2002), offering the possibility of early intervention to ameliorate or compensate for such difficulties. So development the battery for children and adolescents is necessary. Isabelle Peretz, Nathalie Gosselin &Yun Nan exploited the Montreal Battery of Evaluation of Musical Abilities (MBEMA) in 2013. The battery, which comprises tests of memory, scale, contour, interval, and rhythm. The researches was used the MBEMA administered to 245 children in Montreal and 91 in Beijing. The results show that the MBEMA are sensitive to individual differences and to musical training. The abbreviated version was simplified by collapsing the three melodic tests (scale, contour and interval) into a single melody test, and the rhythm and memory tests are identical to the corresponding tests of the full MBEMA. And the abbreviated version was administered to an additional 85 children in Montreal. Their performance indicated that the abbreviated MBEMA are sensitive to individual differences and to musical training too. Thus, the abbreviated MBEMA can serve as an objective, short, and up-to-date test of musical abilities in a variety of situations, from the identification of children with musical difficulties to the assessment of the effects of musical training in typically developing children and perhaps even adults (Isabelle Peretz,Nathalie Gosselin&Yun Nan, 2013). The music educators want to distinguish the congenital amusia is difficult, but the MBEMA can proves a effective tool. It can serve as an effective tool of preliminary screening, if students scored lower on abbreviated version, you can consider to carry on the full MBEMA.

However the Montreal Battery of Evaluation of Musical Abilities (MBEMA) and an abbreviated version have not tested in teenagers, so we not sure whether the abbreviated MBEMA is suitable for Chinese teenagers. The present study sought to verity the stability and effectiveness of the abbreviated MBEMA in a sample of 76 adolescents aged 13-15 years in the city of Tangshan, Hebei province, China.

2. AIMS

The present study sought to verify whether the the abbreviated MBEMA can identification of children with musical difficulties, and assessment of the effects of musical training in 13-15 years old teenagers in China.

3. METHOD

Participants: The sample consisted of 76 teenagers at 13-15 years old form middle schools in Tangshan. Exclusion criteria include brain trauma, hearing deficits, attentional deficits and learning disabilities as reported by the teacher. Most children were right-handed (92.1%) and had no private music lessons (85.5%). Their demographic characteristics are summarized in Table 1.

Table 1:Characteristics of the children tested on the abbreviated MBEMA

Age	13	14	15
Total	41	31	6
Music training			
No lesson	35	26	6
Lessons	6	5	0

The test battery: The abbreviated MBEMA contains scale, rhythm, and memory tests. All tests use the unfamiliar tonal melodies, in 10 different keys. The melodies are computer-generated, each presented in a different timbre. Ten different timbres (e.g., piano, marimba, guitar, flute) are used to make the tests as engaging as possible. Three sets of stimuli, each of which comprising 2 practice trials and 20 test trials, are constructed from these melodies. Each trial consists of a target melody and comparison melody separated by a 1.5-s silent interval; trials are separated by 4-s silent intervals.On each trial, the participants required judge on whether the target and comparison sequence are the same or different. The scale test was first, which has 10 trials with identical (same) comparison melodies and 10 trials with scale-violating (different) comparison melodies. Melody pairs are always presented with the same timbre but in a random order in each test. The rhythm test uses the same melodies as the scale test. Here, the test alter the rhythmic grouping of notes while retaining the number of notes and original meter. The two tests requires a same-different response. The final test of the music battery assesses recognition memory. 10 melodies are selected form the previous tests. Each is presented previously at least four times in the same format. And there are 10 new melodies. The 10 new melodies are constructed in the same manner as the old melodies but differ in their exact temporal and pitch patterns. Participants are required to respond yes if they recognize a melody from earlier in the session and to respond no otherwise. This test assesses incidental memory because participants are not informed that their memory for the melodies will be tested subsequently.

Procedure: The three tests were presented to each participant in a quiet room in their school in a session that lasted 20--30min. The order of presentation was fixed, with scale tests followed by the rhythm and memory tests. The tests were presented through computer loudspeakers at a comfortable listening level. They were free to request breaks between tests.

4. RESULT

The teenagers accuracy (across the three tests) ranged from 43.3% to 96.7% with 50% representing chance performance and 100%, a perfect score. Three students obtained a perfect score in scale test; fourteen students obtained a perfect score in rhythm test; one student obtained a perfect score in memory test; but no one achieved a perfect score in all three tests. As can be seen in Table 2.

Table 2: The average scores of participants

Age	Total	Scale test/20	Rhythm test/20	Memory test/20	Total score
13	41	16.7	17.5	16.05	83.7
14	31	16.4	17.4	16.2	83.1
15	6	18	15.8	16	83

Congenital amusia is typically diagnosed in adults by a global score of 2SD below the mean. Using the same criterion, 3 students out of the 76 tested would be considered amusia. 2 at age 13, 1 at age 14; one teenager has extra curricular music lessons almost 60 month. There scores below the average scores. 1 person present poor performance(55% accuracy rate) in rhythm and memory tests; 2 persons present poor performance in scale test (Table 3). In the following we will eliminate the 3 persons' data.

Table 3:Three amusias situation. In gender"0"represent female;In extracurricular music lesson"1"represent has lessons ; "0" represent has no lesson

Gender	Age	Extracurricular music lesson	Scale test(\20)	Rhythm test(\20)	Memory test（\20)	Total score	Percent correct%
0	14	0	7	13	12	32	46.7
0	13	0	14	11	12	37	61.7
0	13	1	7	15	15	37	61.7

We measured age in terms of months, and interplayed it with the total scores did not reach significance ($F=1.483$， $p=0.234$) and the gender did not reach significance too ($F=0.009$， $p=0.926$). We used independent-sample T test find that extra curricular music lesson with the total scores did not reach significance($F=0.185$， $p=0.668$).However the students who accepted extra curricular music lessons have a better average score in subtest than others(Table 4,in the appendix).

The studied used ANOVA find that there was interaction between age and rhythm test ($F=4.435$,$p=0.015$). We used independent-sample T test find that there was no interaction between extracurricular music lesson and subtest, but there was interaction between gender and scale test ($F=4.434$ ，$p=0.039$).

5. DISCUSSION

None of students obtained a perfect score in all tests, it's confirming the sensitivity of the abbreviated battery. The total score is more sensitive than individual test scores. In the three subtests, the score of memory test was lower than others, because participants are required to respond yes if they recognize a melody from earlier in the session and to respond no otherwise. So this test assesses involuntary memory because participants are not informed that their memory for the melodies will be tested subsequently. Thus, incidental memory of students was worse than the scale and rhythm of the music.

The results show that there was no interaction between age and accuracy it means the total number of correct does not increase or decrease with the growth of the age.The result of rhythm test shows that there was difference between age and accuracy. The older age, the lower average results. The reason maybe that: on the one hand, only one grade participant is in the battery. No stratified sampling on age is one of the shortcomings of this study. On the other hand, the abbreviated MBEMA is much easier for teenagers, which may attribute to the mature abilities of tonal and rhythmic perception at their age. The age has apparent influence on the melody test, the boys results were better than girls in melody test.

The result shows that there was no significant effect between extracurricular music education and test results. However the study (Isabelle Peretz, Nathalie Gosselin &Yun Nan, 2013) shows that months of music lessons predicted total performance at 7 and 8 years, respectively. We analyze the reason may be that in the smaller cities of China, the teachers of extra curricular music education have the limited faculty and level to teach. Even some students did have the extra music education, but non-significant effort in the tests. Besides, as I mentioned earlier, the abbreviated MBEMA is much

easier for teenagers, so whether having received music education, has no influence on the result of the test. But the average score of each subtest results show that the average of participants who accepted the extracurricular music education is better than others. Therefore, the extracurricular music education for teenagers has the effect of music ability, but for some reason, the effect of extracurricular music education is not particularly significant. The results show that the battery on evaluating the training effect of music of Chinese teenagers did not achieve the desired level.

Finally, there are 3 students who scored at or below 2SD from the mean, about 4% of the total number of the participants, that is consistent with previous research results. Thus, the abbreviated MBEMA can be used as a tool for diagnosing deficits in musical skills in adolescents of China. Two students had a melodic difficulty the profile is consistent with the typical profile of amusia adults. And one amusia individual has extracurricular music lessons almost 60 month, but she had a typical melodic deficit, so the amusia is congenital neurologic abnormalities, the music education has no significant for amusia. The 3 students who were tested the full MBEMA in separate sessions 50 days apart. 2 persons would be considered as amusic.

6. CONCLUSIONS

Based on the data analysis, the abbreviated MBEMA can identify effectively that teenagers who were judged as musical disorders (3 out of 76, approximately 4%), and consistent with previous researches. Then the teenagers who might have musical disorders were retested by the full MBEMA. The result shows that they have been diagnosed with Congenital Amusia. Then we can take some measures to guide the teenagers to learning music in right way.

However, there is no significant effect of musical training. This does not accord with previous study (Isabelle Peretz, Nathalie Gosselin &Yun Nan, 2013). Maybe the reason is that the battery is too easier for teenagers. The results show that there was no interaction between age and the total number of correct, only the rhythm test result shows that there was significant effect between age and the total number of correct of rhythm test. That also different from previous study (Isabelle Peretz, Nathalie Gosselin &Yun Nan，2013).

So the abbreviated MBEMA can be used as a tool for diagnosing deficits in musical skills in children and adolescents of China; but it might not be an effective assessment of musical training for 13-15 years old Chinese adolescents. So we must improvement the level of difficulty. Hope we can compiled the diagnostic tests to be more suitable for Chinese youth amusia.

7. REFERENCES

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia:A group study of adults afflicted with a music-specific disorder. Brain, 125, 238−251.

Hyde K, Peretz I. Brains that are out of tune but in time. Psychol Sci 2004;15:356-360.

Hyde, K.L., R.J. Zatorre,et al. 2006. Morphometry of the amusic brain: a two-site study. Brain129:2562–2570（2005）.

Isabelle Peretz,Nathalie Gosselin&Yun Nan. A novel tool for evaluating children's musical abilities across age and culture.Frontiers in systems neuroscience.July 2013 ;Volume 7 Article 30

Isabelle Peretz,Anne Sophie Champod.Varieties of Musical Disorders :The Montreal Battery of Evaluation of Amusia.Ann.N.Y. Acad.Sci.999: 58-75(2003).

Molly J. HenryJ. Devin McAuley.Failure to Apply Signal Detection Theory to the Montreal Battery of Evaluation of Amusia May Misdiagnose Amusia. An Interdisciplinary Journal, Vol. 30, No. 5 (June 2013), pp. 480-496.

Molly J. Henry, J. Devin McAuley.On the Prevalence of Congenital Amusia. An Interdisciplinary Journal, Vol. 27, No. 5 (June 2010), pp. 413-418

Nathalie Gosselin, Pierre Jolicœur, and Isabelle Peretz. Impaired Memory for Pitch in Congenital Amusia. Ann. N.Y. Acad. Sci. 1169: 270–272

Marília Nunes-Silva, Vitor Geraldi Haase.Montreal Battery of Evaluation of Amusia: Validity evidence and norms for adolescents in Belo Horizonte,

Minas Gerais, Brazil.Dement Neuropsychol 2012. December;6(4):244-252

Peretz I, Cummings S, Dubé MP. The genetics of congenital Amusia (Tone Deafness): A family-aggregation study. Am J Hum Genet. 2007; 81:582-588.

Peretz, Isabelle; Gosselin, Nathalie; Tillmann, Barbara.On-line Identification of Congenital Amusia. Music Perception 25.4 (Apr 2008): 331-343.

APPENDIX

Table 4 :The compared between accepted the extracurricular music education subjects and others "1"represent has lessons ; "0" represent has no lesson

Description									
		N	Average	Standard Deviation	Standard Error	The 95% confidence interval of the mean		Minimal value	Maximal value
						lower limit	Upper limit		
Scale test	0	64	16.4375	2.43568	.30446	15.8291	17.0459	7.00	19.00
	1	12	17.6667	1.87487	.54123	16.4754	18.8579	14.00	20.00
	total	76	16.6316	2.38798	.27392	16.0859	17.1773	7.00	20.00
Rhythm test	0	64	17.0313	2.55709	.31964	16.3925	17.6700	7.00	20.00
	1	12	18.0833	2.57464	.74324	16.4475	19.7192	11.00	20.00
	total	76	17.1974	2.57174	.29500	16.6097	17.7850	7.00	20.00
Memory test	0	64	15.9063	2.38859	.29857	15.3096	16.5029	6.00	20.00
	1	12	16.3333	1.92275	.55505	15.1117	17.5550	12.00	18.00
	total	76	15.9737	2.31502	.26555	15.4447	16.5027	6.00	20.00

[2-13] FACTORS INFLUENCING JAPANESE PRESCHOOLERS' INVOLVEMENT IN MUSICAL PLAY AT HOME

Xinyang Huo[1], Mayumi Adachi[2], Emiko Kusanagi[3], Nobuko Hoshi[4], Shing-Jen Chen[5], Tadashi Oishi[6], Hitoshi Takamura[7]
[1]Hokkaido University, Japan, [2]Hokkaido University, Japan, [3]Kokugakuin University Hokkaidou Junior College, Japan, [4]Junior College of Sapporo Otani University, Japan, [5]Koen Gakuen Women's Junior College, Japan, [6]Nara Saho College, Japan, [7]Nara Women's University, Japan
Poster

Preschoolers' musical play has been investigated at playground, daycare, and home. However, it is not clear what determines preschoolers' engagement in a particular musical play. The purpose of this study was to explore factors that would predict Japanese preschoolers' musical play. One hundred and eighteen Japanese mothers with 3- to 6-year-old children, living in different parts of Japan, filled out questionnaires regarding the frequency of various kinds of plays their children were engaged in at home. The questionnaires also included information of how early particular behaviors were observed in children, children's involvement in extra-curricular activities, the mothers' involvement in extra-curricular activities in the past, the frequency of various kinds of maternal behaviors toward children during pregnancy and during one month before survey, the parents' educational background, family income, and children's age and sex. After eliminating samples with incomplete questionnaires, the final sample of the present analyses was 100. To identify predictor variables for the frequency of children's musical play including singing, dancing, and playing a toy instrument, we first conducted correlation analyses between children's involvement in musical play and all of the other variables. Then, we conducted a multiple regression analysis (with backward elimination) with the variables showing significant

correlations with the frequency of children's musical play. The result of the multiple regression analysis showed that children's involvement in musical play at home was predicted by three factors, two of which were the frequency of their involvement in other plays at home (jumping over an elastic rope, $\beta = .28$, VIF = 1.09; pretend play, $\beta = .20$, VIF = 1.13) and the third one was how early children began putting a toy in and out of a box ($\beta = -.25$, VIF = 1.03), $F(3, 96) = 10.23$, $p < .0001$, R^2 (with correction) = .22. This appears to indicate that children's involvement in musical play is influenced by other types of plays, which may also be influenced by developmental and environmental factors. For our presentation, we will explore detailed path analyses to reveal what determines individual differences in preschoolers' musical play at home.

[2-14] AN EXPLORATORY STUDY ON THE FETAL MOVEMENT TO MUSIC

Lingjing Chen[1], Mayumi Adachi[1], Masae Taga[2], Hisanori Minakami[1], Yasushi Handa[1], Masanori Kaneuchi[1]
[1]Hokkaido University, Japan, [2]Sapporo City University, Japan

Poster

In general, the fetus can be exposed to music in two contexts: when the mother is listening to music through headphones and when music is played through loudspeakers. How does the fetus move in these two contexts? What can influence the fetal movement to music? The purpose of the present study was to explore these questions. Thirty-seven pregnant women with singleton fetuses ranging from 23-36 weeks gestational age participated in this experiment. The fetal movement was monitored and video-recorded by ultrasound before any exposure to music (baseline), during the mother's music listening through headphones ("indirect" exposure), and during the same music being played through loudspeakers while the mother was listening to ocean-wave-like sound ("direct" exposure). Music stimuli consisted of 12 one-minute-long cello solo pieces whose pitch ranges were low enough to be transmitted to the womb without extreme amplification. Musical stimuli were randomly presented. The fetal movement of its arm and leg was coded separately, using the Observer XT 11.5 (Noldus). Inter-rater reliability of five fetuses was high ($k = .67$). We conducted lag-sequential analyses (unit of time = 500 ms) for individual fetuses to determine the degree of contingencies (expressed as Z scores) between the fetal movement and the two types of musical exposures as compared with baseline. Results showed that at least 20% of fetuses moved their legs ($\chi2(1, N = 34) = 10.37$, $p = .001$) and 30% moved their arms ($\chi2(1, N = 37) = 6.13$, $p = .013$) contingently to the indirectly exposed music. Moreover, at least 30% of fetuses moved their legs ($\chi2(1, N = 34) = 7.27$, $p = .007$) and 20% moved their arms ($\chi2(1, N = 37) = 7.36$, $p = .007$) contingently to the directly exposed music. We will analyze and report how gestational age, the arousal level of the fetus (measured by the fetal heart rate), the maternal mood (measured by POMS), and the maternal heart rate (monitored continuously) are related to the degree of contingencies between the fetal movement and the type of musical exposure in our poster.

[2-15] MUSIC COGNITION INTO PRACTICE ROOM; IMPLICATION FOR COLLEGE INSTRUMENTAL TEACHER

Sung Hyun Hwang[1]
[1]The Korean Association of Piano Pedagogy, Korea, Republic of

Poster

ABSTRACT

College students, majoring in instrumental performance, seem that they seldom engage their musical knowledge into performing especially in Asia. This research initiate from the doubts about that students are not sufficiently developing learning skills throughout current instrumental pedagogy. For independent career prospects, students are expected to deal with comprehensive repertoires with qualitative development emerging from quantitative practice. However, they are too overwhelmed by distractions to optimize the gulf between theory and performing. Assuming this phenomenon can be solved, proposing in two different perspectives are followed. The first is an extrinsic aspect and the other is an intrinsic aspect. The first proposed solution is cooperative instrumental lesson between performer-teacher and researcher-teacher. Although there are tremendous researches about performing music, college students are not ready to apply it to instrumental skill by themselves due to their tendency of maintaining previous "behavior learning approach". Some researches have been done about this tendency with statistic data, while other researches support ideal resolutions. I integrate two aspects of researches and propose necessity of cross learning system for college students. Secondly, I insist that metacognitive practice need to be informed precisely during lesson. Based on the idea that musical performance has three primary elements; composition, instrument, and performer, I urge that students need to learn how to cognize three sensations, visual, auditory, and kinesthetic. I schematize each sensation to respective primary elements of musical performance, i.e. visual sense to composition, auditory sense to instrument, and kinesthetic sense to performer. Finally I suggest balanced correlations between elements and sensations for emphasizing metacognitive learning. Consequently, considering undergraduate students are the cornerstones of future musical world, I assert that musicological approach of learning, including metacognitive skill would enhance fundamental instrument technique and would accelerate their transformative process toward being independent musician.

1. ISSUE OF COLLEGE CURRICULUM

Since western music introduced in Korea, curriculum for college of music has been established based on western school curriculum, and later, over hundred years, it has been significantly developed. Recently, however, the growth halt between specialized education and integrated education, as well as between quantitative expansion and qualitative expansion. As a result, recent instrumental teachers rarely reflect contemporary musical research methods or current musical issues for their lesson. Most remarkable issue to focus on in this paper is how this phenomenon effects to students related to their independent learning, and suggest possible way to solve this problem.

Most of college provide sufficient amount of theory class beyond instrumental lesson. Even more, most of schools allow students to transfer credit from other university, if circumstances are not capable. However, as college students become their senior year, they are being skeptical about academic curriculum and about its applicability to their career, especially who pursuing professional performer career after graduation. (Jung & Choi, 2013) In the survey from Jung & Choi's research, students want Music Psychology, Music Education or Applied Music as possible additional curriculums, however it is capable to take those classes during academic year if students realize the necessity previously and seek ways. Jung & Choi conclude that students realize importance of theoretical approach for their independent performing career right before their graduation. Another words, even though students notice importance of recent theory for their performing career, they realize the fact so late. In the meantime, as students become their senior year, they have tendency to depend on self-efficacy, mostly by behavior learning approach, to perform instrument. There is research applying 2x2 achievement goal framework and self regulated learning model to college music students, and it shows that during the college year, students achieve their goal by 'performance-approach' orientation in freshman year and then gradually

change its orientation to ‘mastery-avoidance’ in senior year. (Kim, 2010) This result means that student, in freshman year, practice instrument applying cognitive approach, however, they, in senior year, end up practicing instrument by behavior approach with time consuming. Despite the fact that there is other view that 2x2 achievement goal framework is limited to judge instrumental music education, (Miksza, 2009) data from survey is deemed statistically significant. According to Kim’s statistic data, freshmen students attempt various efficient practice strategy for shortening the practice time, however, senior students rely on repetitive practice instead of seeking help from their instrument teacher or research about one’s problem when they complete the musical piece.

To sum up, despite accumulating knowledge during academic year with enough pool of course, college students wander around between theory and instrumental skill. Furthermore, they loosen ability of applying theory to practice instruments. To optimize the gulf between these two, here, suggest that research based instrumental lesson class beyond lesson from skillful performing-teacher. Providing and informing practical usage of theory by research-teacher students are able to be independent for their career achievement after graduation. Moreover, if recent research areas and trends are informed to student a head of time during the academic year, they can make preparation for successful career achievement beyond expanding instrumental skills.

2. THREE SENSES

Beyond above extrinsic change, another issue to bring out is metacognitive practice skills for student’s independent studies.

Basic elements for typical music performing are composition, instrument and performer. Possibly some other elements can be added such as audience, concert hall and so on, however, minimizing as much as performance, at least, could occur, those three elements remain. One may become suspicious considering orchestral piece with conductor, however, it would be easily admitted if the word 'performer' connotes 'operator of instrument'. Another suspicious could evoke about improvisation, but improvisation can be inferred that the piece composed tiny second before performing by performer. Therefore, definitions of above elements are; a composition, an expected sound created by composer; instruments, mechanisms of producing sound; and performers, persons who operate the mechanism. (**Figure.1.1, 1.2**)

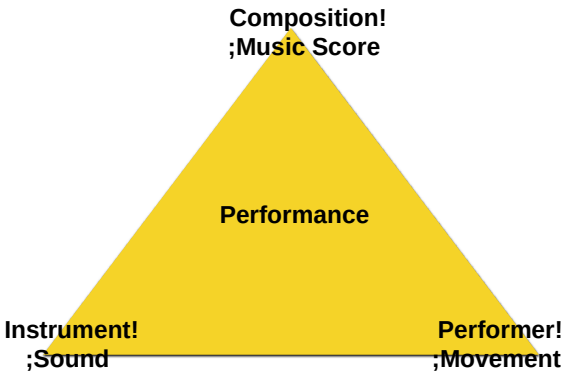


Figure 1.1:Three elements of performance.

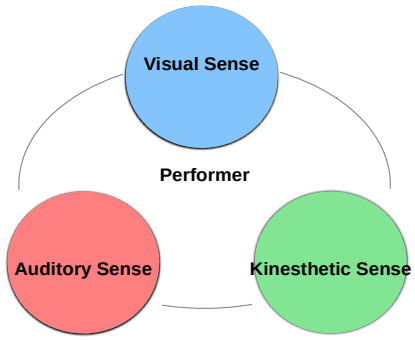


Figure 1.2: Three senses of performer.

Between elements, it is performer that is realistically creating the music performance, because, others are rigid forms. For this reason, instrument possibility is contingent upon performer, and is composition interpretation as well. Since performer is a human being, but not a mechanical object such as computer, expression possibilities counted on one musical artwork are innumerable. These diverse possibilities of musical expression are generated by human potential, and some of this potential are capable to be realized by cognizing sensations with consciousness. Performing music, in some way, is completed by how to cognize three sensations, visual, auditory, and kinesthetic toward primary elements of music, i.e. visual sense to composition, auditory sense to instrument, and kinesthetic sense to performer. Beyond cognizing each sense, balanced correlations between sensations would be the main point for successive expression in music, however, the purpose of this paper is examining and defining the three sensations of performer, a human being. (**Figure 1.3**)



Figure 1.3: Basic correlation between elements and senses (above), and performer’s cognition (below).

2.1. Visual sense

Reading music score means decoding written source and convert it to sound. Visual sense is cognitively applied for decoding musical symbols from the time interpreting it until converting it to intended sound, which means the time before actual sound projected. Envisioned sound is a transitional point of sensing visual sense to auditory sense. Symbolic information in music score is visual form of sound; thereby the symbol does not contain restricted certainties. Regardless the information is notated obviously or ambiguously, easily noticeable or not, every marking is important, because even tiny marking metaphors too many meanings to be neglected. For example, rest of one-sixteenth length at the end of the phrase may effects how to finish this phrase. If one does not perceive this rest and then play previous note unnecessarily long, the end of phrasing would be played far different from composer's intention. Thereby, lack of visual consciousness may lead to fallacious interpretation.

Music score is the most critical foundation source for performer. Reading music score connotes diverse meanings and phases. Generally the implication includes three main points, one is conceiving and interpreting composer's original idea, secondly, is decoding notations and indications in music score itself, and lastly deciding and envisioning when and how the notated symbols are converted to sound (Eggebrecht, 1999). Furthermore, Sloboda (1984)'s experiments show different reading process depending on the level of expert. He found that the most efficient music reading skill is from cognizing the expecting sound of; how loud the sound supposed to be, when exactly the sound projected, and how long the sound supposed to sustain. Also the study found that expert musician receives written information further ahead of time in producing sound, with few mistakes. He concludes that musician, with sufficient time, enable to decode notation in detail exquisitely when they sight read the piece at the first. The more expert musician they are, the more sensitively decode symbols, i.e. modulating metrical information for adding expressive content, considering musical effect more than right notes and right tempo. For expert musician, even beat, bar and pulse on the music score means more than itself. Sloboda said, "Metrical information specified by the score should be specified in the performance too", which means from the reading, musical idea should be conveyed considerably.

Visual sense of written source is not merely enough to be perceived but need to be cognized up to envisioned sound; as if when we read a written letter from someone, we imagine one's pronunciations, intonations and voice of that person. With envisioned sound, notated sources are available to be converted into actual sound.

2.2. Auditory sense

Auditory sense may demand the most delicate cognition among three senses. To satisfy emotional expectation, performer constantly responds to stimulation adjusting both pitch of the sound, and timing of the rhythm. Simultaneously, one operates an instrument as well.

Helmholtz (1998), who investigate physiological theory of acoustics, depicts "Music stands in a much closer connection with pure sensation than any of the other arts", and emphasize that "sensation of tone [sound] are the material of the art [music]". Music, among various arts, is the most impracticable arts form to inquire about perfect definition, because sound, which is the material of music, disappears as soon as it presented. These difficulties arouse inquisitiveness from educator and researcher. There are three general approaches emphasizing importance of auditory sense. Firstly, the approach is accomplished from the view of an educator, secondly from the view of a psychologist, and lastly from a scientist. Educator alludes to necessity of metacognitive skill emphasizing exquisite auditory sense for self-evaluating practice (Hallam, 2001), whereas for the psychologist, rhythm and pitch are frequently issued investigating how emotion is reflected in music(Deutsch, 1982; Krumhansl, 2000). And lastly, the scientist investigates about music as well researching on differences of auditory cortex between musicians and non-musicians (Altenmüller & Wilfried, 2002).

One musical composition has plural interpretations and it is performed differently. Most people accept the idea of that these differences are coming from musician's mind and their musical talent, yet, technically sound is modified by its frequency. .. For instance, the sound is not projected exactly same rhythm and pitch as it is written. According to Krumhansl (2000)'s study, timing of key pressed is not as precise as it is notated. Deutsch (1982), also, describe that sometimes sound is produced small amount ahead of the beat, while sometimes a few milliseconds later than the beat to maintain the tempo depending on the grouped pulse. These delayed or anticipated counting system is a result of auditory stimulation and response. Personal tendency of grouping pitch is the reason why counting system is different respectively depending on performer.

Educators are trying to develop student's auditory sense. However, it is challenging for them, for it is impossible to listen and evaluate the sound on behalf of students all the time. Cognitive learning strategy, which called meta-cognition, seems applicable for expert musicians and is also skeptical depending on students level and development. Despite of another hypothesis that auditory sense is a genetic talent, research on pedagogic methods for elevating auditory sense is actively in progress.

One of the crucial viewpoints here is that, envisioned sound is illusion, but not a practical sound. As a result, auditory sense should activate precisely to ensure corresponding actual sound to envisioned sound.

Projecting sound requires operating instruments. Another words, to facilitate right sound, performer has to move or exercise their body. Some of the body movement requires elaborate executions.

2.3. Kinesthetic sense

Over the last decades, as musical piece has been elevating requisite technique, subjects of kinesthetic sense are researched briskly not only by performing artists but also by physiologists or musicologists. Majorities of performing artists leave the stage in spite of their musical talent when their physical impairments no more retain their instrumental skills. Such things happen when one continuously misuses body without enough understanding of kinesthetic sense and such things are realized after one got injured already. According to the physiological research, if their nerve system does not detect it, people do not feel pain even though they have been gradually injured in muscle or tendon. As a result, when the pain detected by instrumentalists, it is mostly the time when the damage already being progressed far. Moreover, instrumentalists are reluctant to change their posture or way of using finger in sufficient relaxed way innovatively when they start to feel pain, because their muscles are already tamed to play automatically in that way (Mark, 2003). For this reason, it is very important to understand what "relax" exactly mean and to decide which muscle and which finger to use with delicate kinesthetic sensation.

People debate that playing instrument is not sports, however, it is undeniable that instrumentalists are exercising with their instruments behind musical expressions for considerable hours. Practice contains repetitive exercising in order to move finger flawlessly during the performing on the stage, however, if the movement goes ahead of thinking without under control, there would be no spatial for expression.

Kinesthetic approach includes not only general physical skills emphasizing posture of sitting, connection between fingers, hand and forearm, and freedom of movements (Watson, 2009), but also practical technique agonizing how to decide fingering and how haptic sense change tone-colors. Sometimes people are not conscious about what finger they use exactly to press keyboards, especially when they rotate the finger pivoting from thumb. As a result, for instance, even though there are easier way to play scale or arpeggio by substituting another fingering, one keeps practicing repetitively without replacement. Another example of common faults is about using wrist and elbow. Wrist and elbow have more flexibility than what those look like if sensate consciously. With tiny amount of adjust, finger can move faster and more accurate than it with stiffen wrist or elbow. In detail, movements of fingers, hands and arms are related to neck, shoulder and entire torso. For this reason, one need to examine entire body for certain pain, even though it becomes too complicated to seek precise answer (Uszler, et al, 1991, 2000; Mark, 2003).

Another approach of kinesthetic sense, which has been actively investigated, emerges from neuroscience. However most of studies allocate their investigations not for expert instrument performers, but for general public intrigued to music or listener. In addition, neurologically, each sense has intricate and mazelike network for one another, so that it is hard to investigate one sense separately. Kinesthetic sense embodies another two senses, in some way, and is deemed to show varied experimental results depending on circumstances and people, so that it needs further research, especially for performer's actual enigmatic question in mind.

2.4. Summary

To sum up, performer is subject to be studied as valuable as literature researches such as historical composers, compositions analysis and instrumnets mechanism. It seems to have difficulties considering performer is alive and keep changing biologically. However, since musical intelligence are related to other intelligences, various way of interdisciplinary research would be beneficial for both musician and non musician.

3. CONCLUSION

College education for performing instrument aims that students are ready for their professional career and achieve their goal toward music through performing activity. Performing activity is diverse. It not only refers extraordinary solo performers, but also means performers being told to play in any range of conditions. For this reason, the first, instrumental pedagogy should seek creative methods for compelling student to be independent music performer. Secondly, students, themselves, need to enhance both musical and technical skills by seeking their own learning strategies. For both teacher and student, the process of teaching and learning should managed with delicate consciousness from cognizing what they are aim to do with written music, until it is conveyed to intended sound. In this manner, cognitive understanding and intentional behaving are strongly required to be conceived.

4. REFERENCES

Altenmüller, E., & Grun, W. (2002). Brain Mechanisms. In R. Parncutt, & G. E. Mcpherson, The Science & Psychology of Music Performance, *63-81. New York, NY: Oxford University Press.*

Choi, J- h. (2012). Academic Success and Performance Achievement of Music Majors in Korean Universities. Korean Journal of Research in Music Education *, 41 (1), 89-111.*

Clarke, E. (2011). Music Perception and Musical Consciousness. In D. Clarke, & E. Clarke (Eds.), Music and Consciousness: Philosophical, Psychological, and Cultural Perspectives, *193-214. NY, U.S.: Oxford University Press.*

Eggebrecht, H. H. (1999). Understanding Music; The Nature and Limits of Musical Cognition. (R. Evans, Trans.) Burlington, VT: Ashgate Publishing Company.

Deutsch, D. (1982). The Psychology of Music. New York: Academic Press.

Gunther, E., & O'Modhrain, S. (2003). Cutaneous Grooves: Composing for the Sense of Touch. Journal of New Music Research *, 32 (4), 369-381.*

Hallam, S. (2001). The development of metacognition in musicians: Implications for education. British Journal of Music Education *18, 27-39.*

Helmholtz, H. L. (1998). On the Sensation of Tone; as a physiological basis for the theory of music . (A. J. Ellis, Trans.) Bristol, UK: Thoemmes Press.

Jung, H.-h., & Choi, J.-h. (2013). Career Choice and Curriculum Applicability of Music Majors in Korean Universities. Korean Journal of Research in Music Education *, 42 (2), 161-183.*

Kim, J. (2010). A study on the relationship between the 2x2 achievement goal orientation and the self-regulated learning tendency of undergraduate instrumental majors. Graduate Thesis; Ewha Womans University.

Krumhansl, C. L. (2000). Rhythm and Pitch in Music Cognition. Psychological Bulletin *, 126 (1), 159-179.*

Mark, T. (2003). What Every Pianist Needs to Know about the Body. Chicago, IL: GIA Publications, Inc.

Miksza, P. (2009). An investigation of the 2 x 2 achievement goal framework in the context of instrumental music. In L. K. Thompson, & M. R. Campbell (Eds.), Research Perspectives: Thought and Practice in Music Education*. 81-100. U.S.: Information Age Publishing Inc.*

Sloboda, J. A. (1984). Experimental Studies of Music Reading: A Review. Music Perception *, 2 (2), 222-236.*

Uszler, M., Gordon, S., & Smith, S. M. (1991, 2000). The Well-Tempered Keyboard Teacher (2nd Edition ed.). Bermont, CA: Wadsworth, Cengage Learning.

Watson, A. H. (2009). The Biology of Musical Performance and Performance-related Injury. Lanham, MD: Scarecrow Press.

[2-16] APPROACHES ABOUT MOTIVATION FOR THE PRACTICE OF MUSIC IN THE HIGHER EDUCATION: TWO DISCURSIVE POSSIBILITIES

Rosane Cardoso De Araujo[1]
[1]*Federal University of Parana / Cnpq, Brazil*

Poster

This study is on motivation for musical practice in Higher Education based on two theories: Self-efficacy Beliefs and Flow Theory. Self-efficacy Beliefs are the central construct in the Social Cognitive Theory by Albert Bandura. For musicians, the self-efficacy belief affects career perspectives, his behavior towards studying as well as the motivation for musical practice. The second framework is the Flow Theory proposed by Mihaly Csikszentmihalyi that is an approach that has been used especially for the studies focused on the analysis of the quality of the

subject's involvement in the accomplishment of different activities. This study aims is the approach on the issue of motivation to musical practice in the Higher Education ambit, through the intersection of the results of the two researches made in Brazil: a case study about Flow Theory and a survey study about Self-Efficacy Beliefs. The investigated population was students of Music Colleges. In the case study (Flow Theory), by using semi-structured interviews with two musicians, were recognized of the elements of their musical practice related to the flow process: Organization of practice/goals, challenges and repertoire preferences; concentration; emotion and flow. In the survey (Self-Efficacy Beliefs) by using a scale applied to 36 students, the results pointed that the instrumentalist that believes being able to plan his studies may not feel the same way when monitoring his goals and analyzing his performance results. Results of the intersection of the researches, have indicated: (A) that the process of concentration has a relationship with good performance that generates the feeling of satisfaction and, consequently, the persistence in studying; (B) the necessity of the attention to external factors (extrinsic motivation) and internal factors (intrinsic motivation), which guide the actions of the students in their routine music practice; (C) the observation of the confidence that the student has on his performance capacity and the effects of this conduct in the production of results. The principal conclusion pointed was the need of the motivational components of emotion to direct the musical practice, by procedures of regulation (by planning and self-monitoring) and organization (regularity of musical practice), through clear goals.

[2-17] HARMONIC SENSIBILITY OF ELEMENTARY SCHOOL CHILDREN ON CREATING A SUB-MELODY

Junko Takahashi[1]
[1]*Graduate School of Okayama University, Japan*
Poster

The harmonic sensibility is a basic skill, but it has not been regarded as important in Japan. National Curriculum Standards in Japan emphasized functional harmony, and many various studies on harmony, which are often biased towards auditory training. Meanwhile, an international study said children can determine whether progression is good without music education (L.R.Williams, 2005).This study investigated the harmonic sensibility that elementary school children have and how elementary school children have harmonic sensibility when creating a sub-melody. In Experiment 1, The subjects were 4th and 5th grade elementary school children (N=8). The stimuli were 4 melodies. There were 4 measures. The task was creating a sub-melody for each song. The subjects wrote sub-melody in music while listening to the main melody. The wrote melody was evaluated by experts. When only the sub-melody was evaluated, the experts rated them well (M= 3.29, SD=1.20). On the other hand, when the sub-melody was played with the main melody the two were rated low (M= 2.02, SD=1.06). In Experiment 2, how children sing sub-melodies was investigated. Subjects were 5th grade elementary school children (N=35). The stimuli were 2 melodies. There were 4 measures. Task 1 is the famous Japanese song, "Furusato". Task 2 was an original song. After listening to an example sub-melody played over the main melody, the subjects created a sub-melody using an instrument (keyboard, etc.). Subjects took 3-10 days to create new melodies. In task 1, About 30% of the subjects used a sound other than the sound of chord listed in teacher's guide. The intervals in the sub-melody that were used often were: perfect first, perfect fourth and major third. In task 2, the sub-melody created read: C chord, D chord, F chord, G7 chord and Am chord. For a melody they had never heard, the subjects created a melody based on the harmonic sensibility obtained from music learning. Some subjects repeated certain Interval. They could feel a variety of harmonies, and create the sub-melody using the harmonic sensibility that they already had developed from everyday experience.

[2-18] INVESTIGATION OF PERFORMANCE MOTION ON KEYBOARD INSTRUMENT BY DIFFERENCE IN KEY.

Yuki Mito[1], *Hiroshi Kawakami*[1], *Masanobu Miura*[2], *Yukitaka Shinoda*[3]
[1]*Nihon University, College of Art, Japan,* [2]*Ryukoku University, Japan,* [3]*Nihon University, College of Science and Technology, Japan*
Poster

ABSTRACT

We examined the motion of emotional expression of performing a keyboard instrument. We have already examined that of C major, but the motion may vary in different keys. Therefore, in this study, we examined c minor as a different key in comparison with C major. We examined it by using a motion capture system. We calculated center of gravity of the performer's upper body from the motion capture data, and we analyzed whether time series data of the center of gravity changes between different keys. And we compared the difference between the maximum and minimum of the center of gravities of the upper body for each emotion in order to examine difference of keys.

1. INTRODUCTION

We examined the motion of emotional expression of performing a keyboard instrument. In our previous study, we examined the performance motion of five emotions (Juslin, 2001) and emotionless of a professional pianist using a motion capture system (Mito et al, 2013a, 2013b). As a result, we found that obtained motion depends on the emotions, so that we concluded that the motion is influenced by emotional expression. However, the key of the music of the previous study was C major only. The piece of music used for such previous study is shown in Figure 1. We thought that motion may vary in different keys, and we, therefore, made the research for c minor this time because it is a minor key and it begins from the same starting position as same as C major. We compared the results of C major with c minor from these research findings.

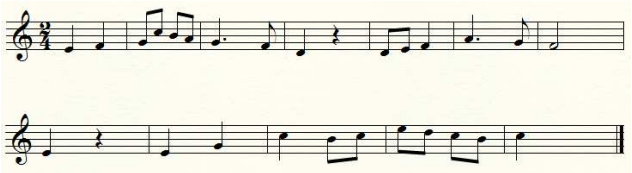


Figure 1 Music of previous study (C major)

2. METHODS

2.1. Performer

We requested a professional pianist to play an electronic keyboard (CASIO CTK-810). Performance motion of him was measured by using the motion capture system.

2.2. Experiment summary

We used a simple piece of music for beginners, which is shown in Figure 2. The key is c minor. The pianist performed this piece by expressing each of five emotions (happiness, tenderness, anger, sadness, fear) used by Juslin or emotionless for the task. He performed the piece with both hands, where right and left hands plays the piece simultaneously with an octave separation. He performed in a suitable tempo after four-beat click counter of M. M. =90.

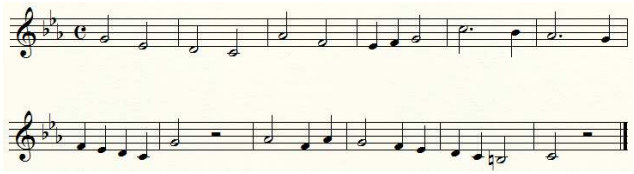


Figure 2 Performance task of this study (c minor)

2.3. Motion measurement

Performance motion was measured by an optical motion capture system (MAC 3D System and Motion Analysis). The reflective markers were attached to the performer's upper body, and the motion was tracked with 6 Raptor-H cameras (frame rate 100 fps, shutter speed 1/2000 s). The measured data were transmitted to a system control computer as two-dimensional data. The data were then analyzed by specific software for motion capture system (Cortex), and changed into three-dimensional data from two-dimensional data in real time. We finally obtained a total of 34 markers on the upper body and keyboard (Figure 3).

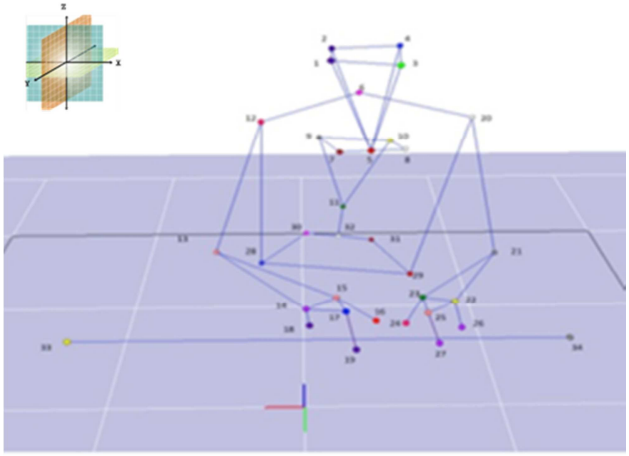


Figure 3 Markers adhesion position

2.4. Analytical approach

The analysis section on the recorded data was from the beginning of the first sound to the last sound. In this study, we tried to compare the motion fluctuation width of the upper body, for which we calculated the center of gravity of the upper body. This was calculated by modeling the upper body as a collection of 8 parts (head, torso, upper arms, forearms, hands), using the center of gravity position of each part $Pg_i(xg(i), yg(i), zg(i))$ (i = 1, 2, ..., 8), the mass center ratio $m(i)$ and the position data of each part of the body obtained from the motion capture data. The center of gravity position $Pg_i(xg(i), yg(i), zg(i))$ of each body part was calculated using Eq. (1). Here, the positions $Ps_i(xs(i), ys(i), zs(i))$ are the start positions of each body part, and the positions $Pe_i(xe(i), ye(i), ze(i))$ are the end positions of each body part.

$$\begin{bmatrix} x_g(i) \\ y_g(i) \\ z_g(i) \end{bmatrix} = \big(1 - m(i)\big) \begin{bmatrix} x_s(i) \\ y_s(i) \\ z_s(i) \end{bmatrix} + m(i) \begin{bmatrix} x_e(i) \\ y_e(i) \\ z_e(i) \end{bmatrix} \quad . \qquad \textbf{(1)}$$

3. RESULTS

3.1. Time series data of the center of gravity

Here we assumed that the upper body of the performer is simply constituted by eight parts (head, torso, upper arms, forearms, hands). The center of gravity of the upper body was calculated from position data and center of gravity ratio. We showed time series data of the center of gravity of the upper body of C major and c minor for each emotion in figure 4-15. Figures were time series data of motion in key array direction of the keyboard instrument (x axis).

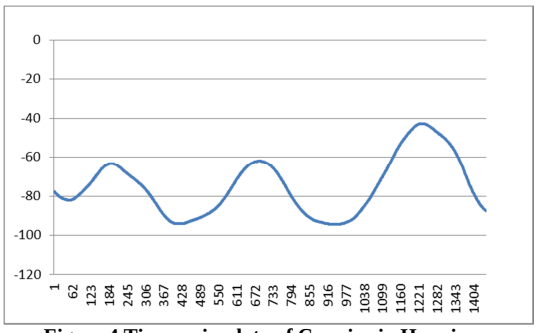


Figure 4 Time series data of C major in Happiness

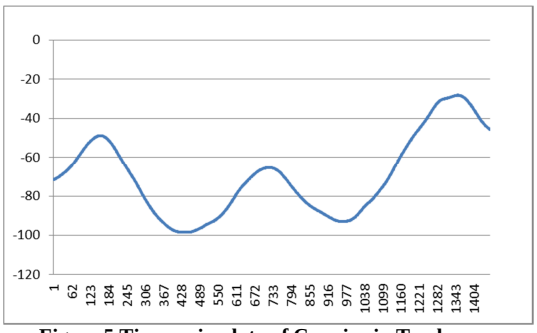


Figure 5 Time series data of C major in Tenderness

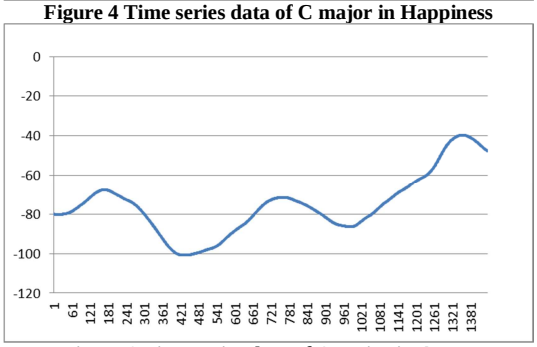


Figure 6 Time series data of C major in Anger

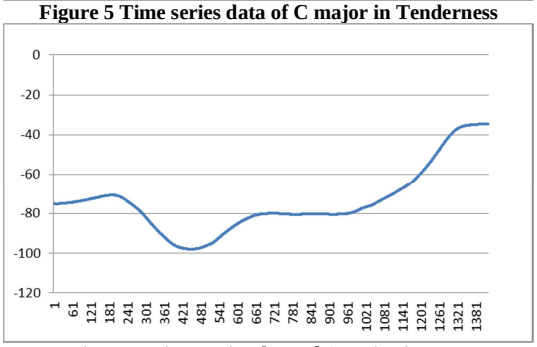


Figure 7 Time series data of C major in Fear

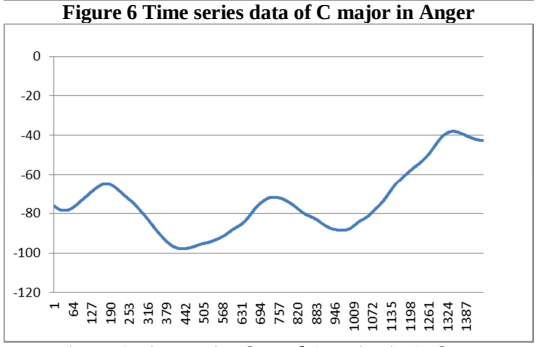


Figure 8 Time series data of C major in Sadness

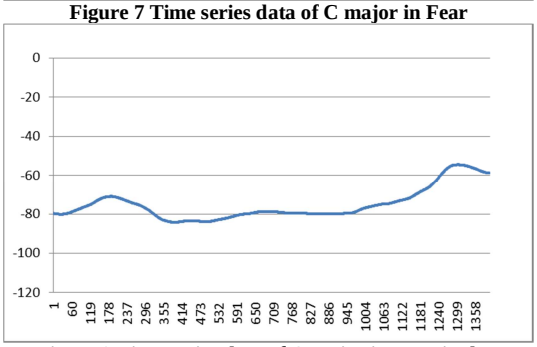


Figure 9 Time series data of C major in Emotionless

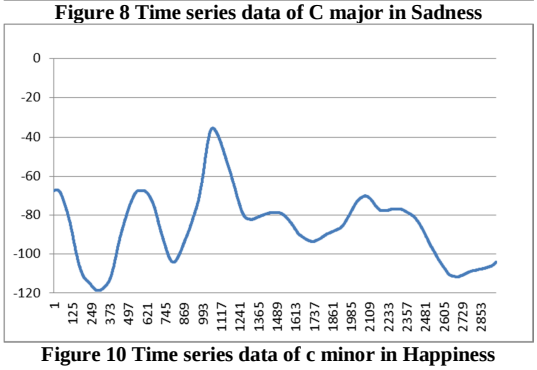


Figure 10 Time series data of c minor in Happiness

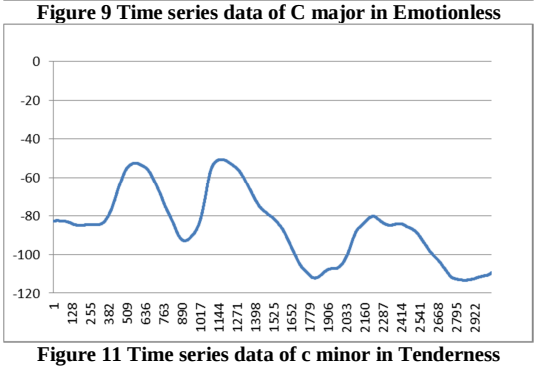


Figure 11 Time series data of c minor in Tenderness

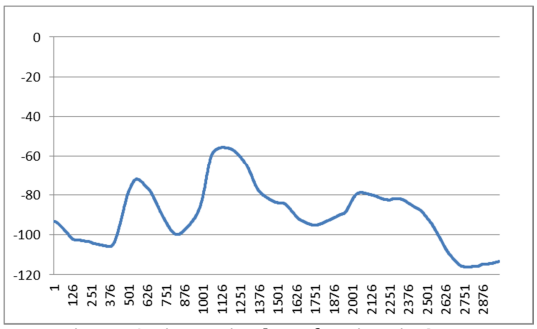


Figure 12 Time series data of c minor in Anger

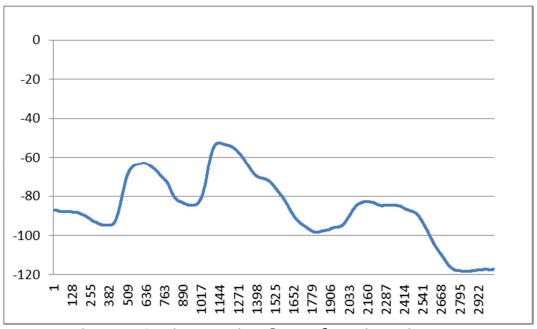


Figure 13 Time series data of c minor in Fear

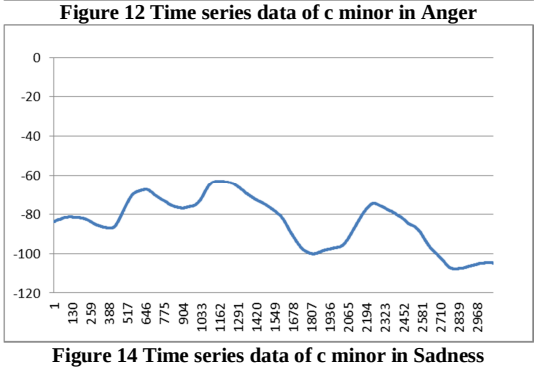


Figure 14 Time series data of c minor in Sadness

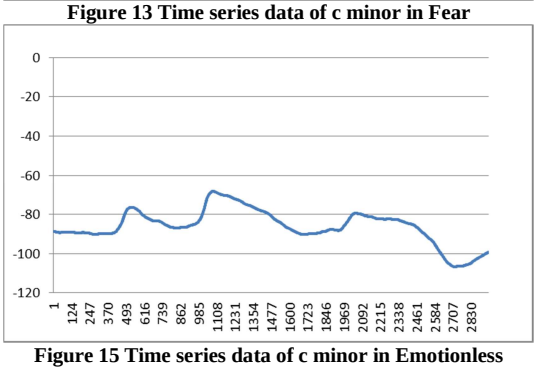


Figure 15 Time series data of c minor in Emotionless

In comparison between positive emotions (Happiness, Tenderness) and negative ones (anger, fear, sadness), positive emotions marked larger fluctuation in both C major and c minor. On the contrary, negative ones had smaller fluctuation. We found that C major and c minor had similar motion data.

3.2. Fluctuation width of center of gravity

We calculated the differences (Δx, Δy, Δz) of each axis between the maximum and minimum positions of the center of gravity of the upper body in three dimension space. The result of C major was shown in Figure 16 and c minor in Figure 17.

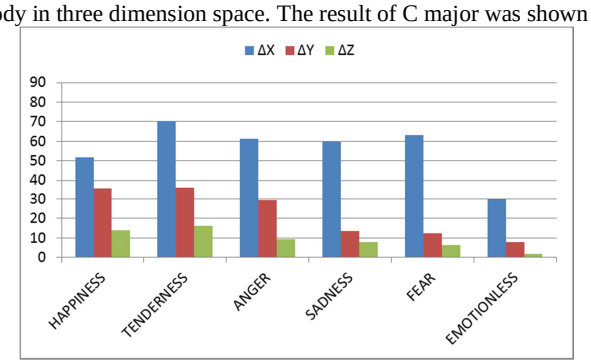


Figure 16 Fluctuation width of center of gravity of upper body in C major

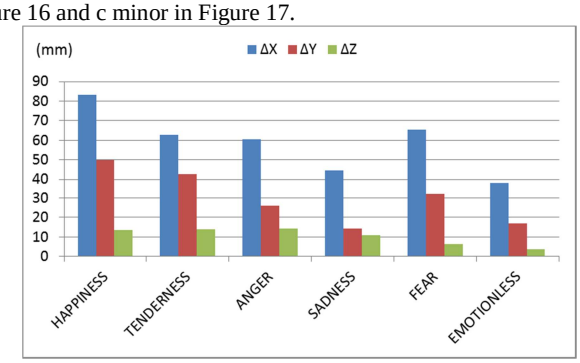


Figure 17 Fluctuation width of center of gravity of upper body in c minor

As a result, the fluctuation width of happiness of C major was smaller. However, that of c minor was larger. Moreover, the fluctuation width of sadness of c minor was smaller, and C major was larger. In the case of performance of happy emotions in C major, exaggerated expression was not found. The reason is that the pianist can play with his usual performance style. In the gloomy emotions such as the sadness, pianist can't express this emotion without playing in an exaggerated style. On the contrary, the happiness emotions in c minor were difficult to express.

In comparison with time series data, we did not find any large change of the performance motion by emotions. The basic emotional expression motion was similar. And the pianist delicately changed expression motion by melody, and changed the image of the music. We found that the pianist slightly changed emotional expression motion by key, and he tried performance to fit his style.

4. CONCLUSION

We investigated the performance motion of performing a keyboard instrument in difference keys, specifically between C major and c minor for five emotions and emotionless. We examined the results by the time series data and fluctuation width of each emotion. In case of the minor key, the motion for happy emotions was large. On the contrary, in case of the major key, the motion for gloomy emotions was large. In this way, we found that the performance motion of keyboard instrument was different by keys.

5. REFERENCES

Juslin, P. (2001). Communicating emotion in music performance: A review and theoretical framework. Music and emotion., Oxford University Press., 309-337

Mito, Y., Kawakami, H., Miura, M. and Shinoda, Y. (2013). The examination of the performance motion and emotional valence by a pianist. Proc. International Congress on Acoustics 2013, POMA Volume 19, 035033

Mito, Y., Kawakami, H., Miura, M., Shinoda, Y. and Marumo, M. (2013). The relation between the key and performance motion on the keyboard instrument. Proc. The International Symposium on Performance Science 2013, 473-478

[2-19] OUTSIDE THE PHONOLOGICAL LOOP? SERIAL ORDER AND SIMILARITY IN WORKING MEMORY FOR TIMBRE

Kai Siedenburg[1], Stephen Mcadams[1], Petr Janata[2]

[1]Centre for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University, Montreal, Qc, Canada, [2]Center For Mind and Brain, Department of Psychology, University of California Davis, Davis, Ca, USA

1. Background: Working memory (WM) denotes the capacity to mentally store, maintain and manipulate limited amounts of information over short periods of time. Research on auditory WM has focused on material that is easily verbalized or sung, such as words or musical pitches. Baddeley's "phonological loop" model posits that auditory WM emerges through the interaction of a buffer and a subvocal rehearsal mechanism. Little is known about how the model extends to material that is difficult to verbalize, such as abstract musical timbres for which semantic labels are not readily available. Based on a pair of experiments using such timbres, an earlier paper (Golubock & Janata, 2013, JEP:HPP), reported low working memory capacity estimates for item identity (< 2 items).

2. Aims: The present report focuses on memory for serial order, that is, the ability of listeners to report the correct position of a probe in the list (in the case of matching "old" items) or the most similar list item (in the case of non-matching "new" items).

3. Methods: We analyzed the data (N=52) from the aforementioned experiment with regard to serial order of timbral sequences that comprised 2, 3, 4, or 6 items and were associated with list–probe delays of 1, 2, 4, or 6 seconds.

4. Results: Using proportion correct responses (corrected for a length-dependent chance baseline) as the dependent variable, repeated-measures ANOVAs showed robust effects of length, delay and serial position. This indicates a limited memory capacity for serial order which is shaped by primacy/recency gradients. We further observed that distance in physical timbre space between the probe and its nearest neighbor in the list strongly correlated with performance in both item and order recognition tasks, suggesting a similarity effect for timbre.

5. Conclusions: These findings structurally resemble hallmark effects of WM data that use verbal materials. Nonetheless, capacity estimates for item identity had fallen under what could have been expected for material bearing semantic connotation and the timbres employed seem unsuited to what is classically described by verbal rehearsal. This challenges the phonological loop as a model of general auditory WM.

[2-20] VOCAL MELODY EXTRACTION BASED ON BAYESIAN FRAMWORK

Liming Song[1], Ming Li[1], Yonghong Yan[1]
[1]Chinese Academy of Sciences, China

Poster

ABSTRACT

We present a Bayesian framework-based approach for automatically extracting the main melody from polyphonic music, especially vocal melody songs. According to various informations from the music signals, we use a pitch evolution model describing how pitch contour changes and an acoustic model representing the acoustic characteristics when the pitch is hypothesized, and search for the best pitch sequence utilizing Viterbi algorithm. The experimental results on the RWC popular music database indicate that the overall accuracy achieves 76.4%. The overall accuracy is 69.2% on MIR-1K dataset at 0 dB mixed level.

1. INTRODUCTION

Melody is the a concise and representative description of polyphonic music defined as a succession of the predominant fundamental frequency of the musical source. It can be used in numerous applications such as automatic music transcription, music structure analysis, Query-by-humming system and music information retrieval. In order to solve this challenge, many melody extraction algorithms have been proposed in the last decade. Goto for the first time used a monophonic fundamental frequency (F0) sequence to represent the melody and achieved automatic transcription task from real world music with his famous PreFEst algorithm (Goto, 2004). Dressler introduced a new approach based on the pair-wise analysis of spectral peaks for identification of partials with successive (odd) harmonic numbers (Dressler, 2011). Most of the melody extraction algorithms attempt to cluster the partial peaks to various different sources, and track the melody from these peaks. The method presented by Cao uses the subharmonic summation spectrum as a salience function of the pitch, and tracks the melodic line with a harmonic structure tracking strategy (Cao, 2007). Recently, some researchers present a novel set of approaches by separating the melody source from the mixture using timbre-based source separation algorithms. Durrieu (Durrieu, 2010) uses two separate timbre models, one for the melody (sometimes specifically human singing voice) and the other for the accompaniment, and reaches a better extraction performance.

However, most researchers noticed the fact that the challenge of tracking the main melody from the real word polyphonic music signal can't be resolved satisfactorily just relying upon the classical signal processing solution. It is necessary to utilize some musicological knowledge or auditory perception theory. The method proposed by Klapuri consists of a computational model of the human auditory periphery, which is good at utilizing the higher order overtones of harmonic sound (Klapuri, 2008). But when analyzing clean, wide-band signals, the auditory model did not have a clear advantage, since most of the energy of music and speech sound is at the lower partials, for which the bandwidth nonlinearity is less important. Thornburg considered the melodic context and modeled it as sequences of note events. In this way, the system readily distinguishes abrupt spectral changes associated with musical onsets from other abrupt change events (Thornburg , 2007).

The approaches mentioned above have utilized the musicological information of the music. However, most of them only consider the transition of adjacent notes, but do not distinguish the pitch lines generated by other accompaniment instruments, such as string instrument. In this paper, we present a method under Bayesian framework, making sufficient use of the melodic context and also the differences of acoustic characteristics between the voice and other instruments. Figure 1 is the block diagram of our system, which consists of two modules: a pitch

evolution model and an acoustic model. Pitch evolution model describes how pitch contours change, which contains two sub-models that are pitch transition model and harmonics variation model. The acoustic model represents what the acoustic characteristics would be when the pitch is the hypothetical one. The acoustic model includes three sub-models that are pitch periodicity model, harmonic shape model and vocal/non-vocal model. The output melody contour is obtained via Viterbi algorithm by maximizing the weighted summation of the scores of each sub-models.

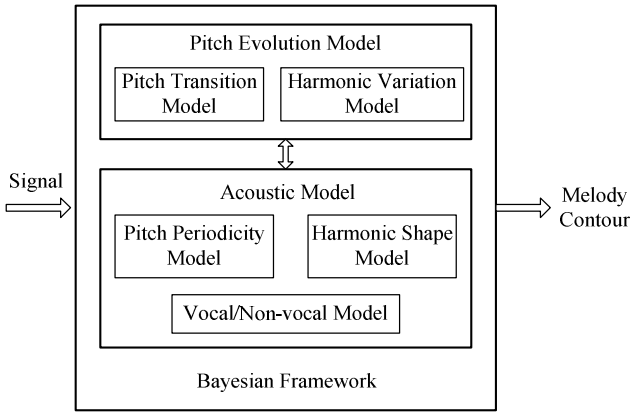


Figure 1 The block diagram of the system

2. METHOD

We assume $\mathbf{F}_0 = \{ f_{0,1}, f_{0,2}, ..., f_{0,t} \}$ is the pitch sequence of a polyphonic music song, $\mathbf{O} = \{O_1, O_2, ..., O_t\}$ is the characteristics observed from the input musical signal, and $\mathbf{F_0^c} = \{ f_{0,1}^c, f_{0,2}^c, ... f_{0,t}^c, \}$ is the pitch candidate series. The optimal pitch sequence satisfies the following equation (1), where Bayes theorem is applied.

$$\begin{aligned} \mathbf{F}_0 &= \arg\max_{\mathbf{F}_0^c} p(\mathbf{F}_0^c \mid \mathbf{O}) \\ &= \arg\max_{\mathbf{F}_0^c} p(\mathbf{F}_0^c)\, p(\mathbf{O} \mid \mathbf{F}_0^c) \end{aligned} \quad (1)$$

Where, $p(\mathbf{F_0^c})$ denotes the probability of the appearance of the potential candidate pitch series. We consider that the probability of the appearance of the pitch $f_{0,t}^c$ only relates to the previous adjacent one $f_{0,t-1}^c$,

$$\begin{aligned} p(\mathbf{F}_0^c) &= p(f_{0,1}^c, f_{0,2}^c, ... f_{0,N}^c) \\ &= p(f_{0,1}^c) \prod_{t=2}^{N} p(f_{0,t}^c \mid f_{0,t-1}^c) \end{aligned} \quad (2)$$

Where N is the number of frames of the music signal.

$p(\mathbf{O} \mid \mathbf{F_0^c})$ denotes the probability of the observation characteristics given a pitch contour candidate.

$$p(\mathbf{O} \mid \mathbf{F}_0^c) = \prod_{t=1}^{N} p(O_t \mid f_{0,t}^c) \quad (3)$$

For simplicity, we assume that the probability of observation characteristics for given pitch candidate of each frame is independent each other.

2.1. Pitch Evolution Model

Considering that the melody contour is continuous in time and changes gradually, as well as the magnitude of the multiples of F0 and the pitch is only associated with the previous situation, the pitch evolution model is defined as

$$\begin{aligned} p(f_{0,t}^c \mid f_{0,t-1}^c) &= p(\Delta f, V_t \mid f_{0,t-1}^c) \\ &= p(\Delta f \mid f_{0,t-1}^c)\, p(V_t \mid f_{0,t-1}^c) \end{aligned} \quad (4)$$

Where $p(\Delta f \mid f_{0,t-1}^c)$ is the pitch transition model representing the change of the pitch from one frame to the next. $p(V_t \mid f_{0,t-1}^c)$ is named as harmonics variation model represents the variation of the magnitude of the partials of $f_{0,t-1}^c$. We assume the two sub-models are independent for each other.

As suggested in (Wu, 2003), the pitch transition $\Delta f = f_{0,t-1}^c - f_{0,t}^c$ can be modeled by a Laplacian distribution.

$$p(\Delta f \mid f_{0,t-1}^{c}) = \frac{1}{2\sigma}\exp(-\frac{|\Delta f| - \mu}{\sigma}) \quad (5)$$

Where μ and σ are bias and spread, respectively. We let μ=0.07, σ=5.2 , and limit $|\Delta f|$ <30Hz since pitch cannot change too fast in short time.

Observing the spectrum of the music signal, we can find that the melody F0 and the partials present as a comb structure. In other words, the magnitude of each harmonic of pitch is usually continuous in time. We use the normalized cross-correlation of the energy vectors $\vec{\mathbf{E}}^{h} = \{X(f_0^c), X(2f_0^c), ..., X(Mf_0^c)\}$ of the harmonics of the neighboring frames to indicate the continuity of harmonic variation. Where M is the maximum number of harmonics, $X(f,t)$ is the magnitude of frequency f at frame t.

$$p(V_t \mid f_{0,t-1}^{c}) = \frac{\vec{\mathbf{E}}_t^{h} \cdot \vec{\mathbf{E}}_{t-1}^{h}}{\left|\vec{\mathbf{E}}_t^{h}\right| \cdot \left|\vec{\mathbf{E}}_{t-1}^{h}\right|} \quad (6)$$

2.2. Acoustic Model

The acoustic model given a pitch is defined as follow:

$$\begin{aligned} p(O_t \mid f_{0,t}^{c}) &= p(E_t^{s}, M_t, S_t \mid f_{0,t}^{c}) \\ &= p(E_t^{s} \mid f_{0,t}^{c})\, p(M_t \mid f_{0,t}^{c})\, p(S_t \mid f_{0,t}^{c}) \end{aligned} \quad (7)$$

Where E_t^{s} represents the pitch periodicity, Mt relates to the harmonic shape, and St denotes whether the vocal exists, i.e., whether melody pitch presents at a certain time frame or not. We assume that the three sub acoustic models are independent each other.

Here we use the harmonic salience summation of the pitch candidate reported in (Song, 2013), concluded as:

$$E_t^{s} = \sum_{m=1}^{M} h^{m-1} \frac{X(mf_0^{c}, t)}{M_{f_0^{c}}(f,t)} \quad (8)$$

Where X(f, t) is the power spectrum value of the frequency f at tth frame, and h is a compression factor, 0<h<1. Mf is an inhibiting factor, which can suppress the values of the spectrum peaks when calculating the harmonic salience of half- and multi- fundamental frequency.

In despite of the string instrument presents a similar structure in the spectrum, but the distribution of the harmonic energy is different from that of voice. The harmonic structure of instruments has a different shape in the spectrum compared with that of voice. The distribution of the harmonic energy of voice is not as uniform as the one of string. We use the harmonic energy vector $\vec{\mathbf{E}}^{h}$ to describe the shape, and Gaussian Mixture Model (GMM) to measure the posterior probability $p(M_t \mid f_{0,t}^{c})$ for the given pitch candidate $f_{0,t}^{c}$ at frame t.

To decide whether the vocal exists at a certain time is a key task in melody extraction, named as Vocal/Non-vocal (VNV) decision. The human voice as a special instrument, has its own unique timbre characteristic compared to other instruments. We use a combination of Shifted Delta Cepstra for Mel-Frequency Cepstrum Coefficients (MFCC), Spectral Contrast Feature (SCF) and Harmonic Features(HF) as the features to train two models via GMM: Vocal model λV and Non-vocal model λNV.

MFCC (Davis, 1980) is cepstral coefficients calculated on a mel-frequency scale, and fitted to the characteristics of the human auditory sense. In order to raise the capability of describing long term dynamic variation of features, Shifted Delta Cepstra algorithm is applied to MFCC, named as MSDC, which is a group of differential features of adjacent features. Spectral Contrast Feature (Jiang, 2002) represents the relative spectral characteristics in each subband which is a good compensation for MFCC that reflects the average spectral characteristics. Harmonic Features contain a set of features related to pitch and harmonics (Geoffroy, 2004) including Inharmonicity, Harmonic Spectral Deviation, Odd to Even Harmonic Energy Ratio, Tristimulus (Pollard, 1982) and Pitch Salience.

Then the VNV likelihood can be measured as

$$S_t = \log p(\mathbf{x} \mid \lambda_V) - \log p(\mathbf{x} \mid \lambda_{NV}) \quad (9)$$

At last, we perform a Viterbi algorithm to search the optimal melodic line. According to the equations (1), (4), (7), the probability of the appearance of the candidate pitch $f_{0,t}^{i}$ is decided by:

$$\begin{aligned} p(f_{0,t}^{i}) = \arg\max_{j} [&\log p(f_{0,t-1}^{j}) + w_f \log p(\Delta f_{j\to i} \mid f_{0,t-1}^{j}) \\ &+ w_v \log p(V_{t,j\to i} \mid f_{0,t-1}^{j})] + w_e \log p(E_t^{s} \mid f_{0,t}^{i}) \\ &+ w_m \log p(M_t \mid f_{0,t}^{i}) + w_s \log p(S_t \mid f_{0,t}^{i}) \end{aligned} \quad (10)$$

Where w* is the weight of each sub-model. $f_{0,t}^{i}$ is the ith candidate pitch at frame t.

3. RESULTS

For evaluation, we chose two databases to test the performance, which are Popular Music Dataset (Goto, 2004, April) and MIR-1K database (Hsu, 2010). RWC database contains 80 Japanese pop songs and 20 English pop songs that were mixed just as a producer originally intended. The recorded sounds of RWC are provided as monaural sound files at 16 bit / 44.1 kHz in the RIFF WAVE format. Each song has been annotated the singing melody by human listeners. The duration of each song ranges from 3 minutes to 6 minutes.

MIR-1K database contains 1000 song clips recorded at 16-kHz sampling rate. The duration of each clip ranges from 4 to 13 seconds and the total length of the dataset is 133 minutes. These clips were extracted from 110 karaoke songs which contain a mixed track and a music accompaniment track. All clips were annotated with information including pitch contour in semitone, indices and types for unvoiced frames. Most of the singers are lab students who are amateurs with no professional training. The music accompaniment and the singing voice were recorded at the left and right channels, respectively. In our experiments we mixed the vocal channel and accompaniment channel at 0 dB SNR.

For each dataset, we conducted a 4-fold cross validation. Each database was divided into four non-overlapping sets. Then we repeated the following step four times: Left out three of the four sets for training and use the rest one for testing. For comparison, the methods proposed by (Cao, 2007), (Ellis, 2006), (Salamon, 2012) and (Klapuri, 2008) were uesd on RWC dataset. Table1 shows the results for the proposed method and for the other comparison methods on the RWC databases. It can be seen that although the recall ratio of the proposed method is lower than that of the others, but the false alarm is much lower, only 12.6%, less than a third of the others. The raw pitch accuracy of our method reaches 77.1%, higher 7.6% than the method presented by C.Cao. As the most important evaluation criterion in melody extraction, the overall accuracy of the proposed method achieves 76.4%, which is 11.8% higher than the other algorithms at least.

Table 1 Melody extraction results in RWC dataset

Method	Vc det(%)	Vc FA(%)	RawP(%)	Allacc(%)
Ellis	93.4	66.3	50.4	42.9
C.Cao	86.26	41.57	69.59	64.53
Salamon	86.65	37.94	65.01	59.21
Klapuri	87.00	39.40	63.60	63.00
Proposed	**78.44**	**12.62**	**77.19**	**76.61**

1 Vc det denotes the recall ratio of voicing detection, while Vc FA the voicing false alarm. RawP represents the raw pitch accuracy of the final pitch, while Allacc the overall accuracy of the melody.

MIR-1K is the main test database of melody extraction task in Music Information Retrieval Evaluation eXchange (MIREX) in recent years. Table2 shows the contest results in MIREX 2012 and MIREX 2013 at 0 dB and the results of our algorithm. Though Yeh method gives highest RawP, it does not detect vocal/non-vocal parts at all, which leads to a much lower overall accuracy. Myer method has a higher voicing false alarm that makes the overall accuracy a little lower. Our proposed method is comparable with Arora approach.

Table 2 Melody Extraction Evaluation in MIREX 2012~2013

	Method	Vc det(%)	Vc FA(%)	RawP(%)	Allacc(%)
2012	Myer	81.80	37.85	70.32	65.57
	Arora	79.24	22.48	69.55	68.84
	Yeh1	100.00	100.00	83.87	54.19
	Yeh2	100.00	99.78	80.19	51.77
2013	Song	87.51	33.75	63.96	63.88
	Canna	27.01	8.94	43.87	46.21
	Yeh3	100.0	99.78	80.19	51.77
	Yeh4	100.0	100.0	83.87	54.19
Proposed		**85.07**	**22.40**	**67.57**	**69.2**

4. CONCLUSION

We proposed an approach for the automatic extraction of melody in polyphonic music. The method consists of a pitch evolution model and an acoustic model, making sufficient use of the melodic context and also the differences of acoustic characteristics between the voice and other instruments, and uses the Viterbi algorithm to find the optimal melodic line.

The improvement of the overall accuracy is 11.8% reaching 76.4% in RWC dataset. It has the best performance in overall accuracy compared with the algorithms in melody extraction evaluation of MIREX 2012~2013.

5. R EFERENCES

Cao, C., Li, M., Liu, J., & Yan, Y. (2007, September). Singing Melody Extraction in Polyphonic Music by Harmonic Tracking. In ISMIR *(pp. 373-374).*

Davis, Steven and Mermelstein, Paul(1980), "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," Acoustics, Speech and Signal Processing, IEEE Transactions on, Vol.~28, No.~4, pp.~357--366, 1980.

Dressler, K. (2011, July). Pitch estimation by the pair-wise evaluation of spectral peaks. In Audio Engineering Society Conference: 42nd International Conference: Semantic Audio. *Audio Engineering Society.*

Durrieu, J. L., Richard, G., David, B., & Févotte, C. (2010). Source/filter model for unsupervised main melody extraction from polyphonic audio signals. Audio, Speech, and Language Processing, IEEE Transactions on,18*(3), 564-575.*

Ellis, D. P., & Poliner, G. E. (2006). Classification-based melody transcription. Machine Learning, 65(2-3), *439-456.*

Goto, M. (2004, April). Development of the RWC music database. InProceedings of the 18th International Congress on Acoustics (ICA 2004), Vol. 1, pp. 553-556.

*Goto, M. (2004). A real-time music-scene-description system: predominant-f0 estimation for detecting melody and bass lines in real-world audio signals.*Speech Communication, 43*(4), 311-329.*

Geoffroy Peeters(2004), "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," 2004.

Hsu C L, Jang J S R(2010), "On the improvement of singing voice separation for monaural recordings using the MIR-1K dataset," Audio, Speech, and Language Processing, IEEE Transactions on,Vol.~18, No.~2, pp.~310-319, 2010.

Jiang Dan-Ning, Lie Lu, Hong-Jiang Zhang, Jian-Hua Tao, and Lian-Hong Cai(2002). "Music type classification by spectral contrast feature," In Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on, *volume 1, pages 113–116. IEEE, 2002.*

Jin, Z., & Wang, D. (2011). HMM-based multipitch tracking for noisy and reverberant speech. Audio, Speech, and Language Processing, IEEE Transactions on, 19(5), *1091-1102.*

Klapuri, A. (2008). Multipitch analysis of polyphonic music and speech signals using an auditory model. Audio, Speech, and Language Processing, IEEE Transactions on, 16(2), *255-266.*

Nwe, T. L., & Wang, Y. (2004, October). Automatic Detection Of Vocal Segments In Popular Songs. In ISMIR.

Salamon, J., & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. Audio, Speech, and Language Processing, IEEE Transactions on, 20(6), *1759-1770.*

Song, L., Li, M., & Yan, Y. (2013). Pitch estimation based on harmonic salience. Electronics Letters, 49(23), *1491-1492.*

Thornburg, H., Leistikow, R. J., & Berger, J. (2007). Melody extraction and musical onset detection via probabilistic models of framewise STFT peak data. Audio, Speech, and Language Processing, IEEE Transactions on,15(4), *1257-1272.*

Pollard, H. F., & Jansson, E. V. (1982). A tristimulus method for the specification of musical timbre. Acta Acustica united with Acustica, 51(3), *162-171.*

Wu, M., Wang, D., & Brown, G. J. (2003). A multipitch tracking algorithm for noisy speech. Speech and Audio Processing, IEEE Transactions on, 11(3), *229-241.*

[2-21] EFFECTS OF TONAL CONTEXT ON THE PERCEPTION OF PITCH INTERVALS

Jackson Graves[1], Andrew Oxenham[1]
[1]University of Minnesota, USA

Poster

The auditory perceptual dimension of pitch carries essential information for perceiving speech and music. Structures of related pitches such as scales, keys, and local harmonies form tonal hierarchies, which may be an important contributor to the brain's impressive capacity for processing patterns of musical pitch, such as melodies. To test the hypothesis that tonal context allows melodies to be processed as points in a tonal structure rather than a series of intervals, listeners performed a pitch interval discrimination task in five context conditions: no context, repetition, mistuned, whole tone, and major scale. Context was found to significantly affect performance on this task, revealing that intervals are more reliably discriminated when preceded by tone sequences that establish a musical key. The results suggest that melody perception may be coded more efficiently through longer-term relations within a tonal hierarchy than through a simple encoding of successive interval sizes.

[2-22] INTERVAL AND BEAT-BASED STRATEGIES OF RHYTHM PERCEPTION IN MUSICIANS AND NON-MUSICIANS

Rafał Lawendowski[1], Krzysztof Basiński[1]
[1]University of Gdańsk, Institute of Psychology, Poland

Poster

ABSTRACT

The ability to perceive rhythm is necessary for activities such as dancing or playing music. Many models of rhythm perception have been proposed throughout the history of cognitive science. These models fall into two main theoretical frameworks: an interval-based or entrainment-based perspective. Both these models have not been successful in explaining all aspects of rhythm perception. Thus, some authors suggest that humans may use different strategies when perceiving rhythmical stimuli, listening in interval or entrainment (beat) mode. It is proposed that there are individual differences in the tendency to prefer one mode or the other. The aim of the current study was to examine if subjects with formal musical education have different strategies of rhythm perception than subjects with no musical education whatsoever. We hypothesized that trained musicians will tend to perceive rhythmical stimuli using beat-based strategy because of their experience with tasks associated with musical rhythm and tempo judgment.

Non-musician (N=25) and musician (N=22) groups were recruited for the study. To assess if subjects perceive rhythm in an interval or beat mode, we used an experimental procedure developed by McAuley and colleagues. This method involves listening to sequences of rhythmically presented tones with last tone shifted forwards or backwards in time. The participants were asked to judge if the presented sequences felt like 'slowing down' or 'speeding up'. Analysis showed no relationship between rhythm perception strategies and formal musical education. Participants assessed the stimuli using two different strategies regardless of their prior musical education. Other potential sources of this variability are explored, including genetic and biological factors, dependence on context and a possible lack of stability of individual differences in rhythm perception strategies over time.

1. INTRODUCTION

Perception is a cognitive process, within which, humans systematize knowledge concerning the surrounding world. Perception may be associated with the way in which humans organize musical information. Repp (2007) states that musical sequences can be considered opaque figures that individuals can perceptually organize in different ways.

Rhythm perception is a notion of great importance as the rhythmic structure itself is an indispensable and fundamental ingredient of music, regardless of its style or genre. Rhythm, which is an integral part of music, is also a structural element of our environment. Quantum physics states that every atom pulsates about a million billion times a second, while the Earth vibrates once every 53.1 minutes (Chen, 1983). Humans have been living in such rhythmic environment for ages, the environment which regular rhythm is determined by seasons, moon phases, days being followed by nights. That is why it is widely understood that rhythmic activity seems to be universal for all the people and that predispositions to it are even visible in our pre-natal life. Its presence at this early stage of life proves that rhythmic activity is evolutionary conditioned (Honing, Ploeger, 2012). Our organisms are defined by such rhythmic patterns as heart beat, breath or brain waves.

Music, with the usage of the system of motor control, makes people dance, hop, clap or snap fingers in time. The ability to adjust our bodies and keep time gives great pleasure. One of the reasons why human beings are especially sensitive to rhythm is the fact that the first acoustic stimulus we experience in mother's womb is the rhythm of her heartbeat. After being born, child is still strongly influenced by his/her mother's heart. It stems from the fact that about 80% of mothers hold their children on the left arm (Harrer, 1975).

Rhythm is also an integral part of the language (Patel, 2007). Without it, communication or any other social interactions could be hindered, or even impossible. In the continuous speech stream, syllable boundaries are marked by slow rhythmic modulations at ~1- 8 Hz (Hickok, Poeppel, 2004). Perception of syllable rhythm is a crucial gateway for phonological representation establishment. Infants/children are sensitive to speech rhythm variations. They use prosodic cues to tune into syntax and learn vocabulary (Fernald & McRoberts, 1996; Soderstrom et al, 2003). What is interesting, dysrhythmia may indicate certain disturbances, as it is visible in the case of dyslexia (Overy, 2000). Children with dyslexia show deficits in temporal processing, both in language and in music. Deficit in rhythmic sensitivity is a core deficit in dyslexia (Goswami, Thomson et al, 2002). Overall, the way in which the listener organizes, analyses and interprets rhythm is one of the most important aspects in human psychological activity.

2. INTERVAL AND ENTRAINMENT-BASED MODELS OF RHYTHM PERCEPTION

Literature on rhythm perception involves many different approaches and paradigms, from research on neural mechanisms underlying timing in humans (Grahn, 2012), to various studies of non-human vocal-learning species (Fitch, 2006; Patel, Iversen, Bregman, Schulz, & Schulz, 2008). Many researchers have proposed different theoretical models of rhythm perception and sequence timing in general. These models can be categorized into two main theoretical frameworks: the interval-based and entrainment-based views (McAuley, Frater, Janke, & Miller, 2006; McAuley, 2010).

Historically first was the interval-based approach to timing, with roots in information-processing framework. One of the most regarded and influential theories in this approach is the Scalar Expectancy Theory proposed by Gibbon (1977). It divides temporal processing into three distinct stages: the clock, memory and decision stage. The clock consists of an internal pacemaker, an oscillator that emits regular pulses. These pulses flow into the accumulator via a switch controlled by attention. The switch opens and closes at the start and stop times of the incoming stimulus. The memory stage involves a mechanism of evaluating the number of pulses present in the accumulator and storing them in long-term reference memory. This data is then compared during the decision stage to form a temporal judgment, such as "shorter" or "longer" (Gibbon, 1977; McAuley, 2010).

The entrainment-based perspective is a newer theoretical framework based on dynamic systems perspective. Entrainment-based theories propose that humans have the ability to synchronize to rhythms and tempos present in the environment. The basis of this perspective is the observation of circadian rhythms, such as the sleep-wake cycle. The environmental rhythm of day-night cycles (the so called *driving* rhythm) entrains the biological rhythm of organisms (the *driven* rhythm) represented in the periodic activity of the nervous system. When adapted to music (or any rhythm with periodicity ranging from a few hundred milliseconds to a few seconds), this approach similarly addresses the concept of entrainable oscillation. Judgments about the temporal structure of rhythms are based on synchronizing (*entraining*) the internal oscillator to a given musical rhythm (Grahn, 2012; Large & Jones, 1999; McAuley, 2010).

3. INDIVIDUAL DIFFERENCES IN RHYTHM PERCEPTION

Although these two frameworks of rhythm perception seem contradictory, some researchers suggest that humans may in fact perceive rhythm using two different strategies based on two aforementioned models. In this view, the tendency to prefer one strategy over the other is attributed to individual differences in the mode of listening (McAuley et al., 2006). This idea is based on the fact that both theoretical models have met with mixed success in describing and predicting the outcomes of the whole spectrum of rhythm tasks. Interval-based approaches achieve better results in single interval judgment tasks, while entrainment-based approaches are more successful in predicting outcomes of sequence timing tasks (Keele, Nicoletti, Ivry, & Pokorny, 1989; McAuley & Jones, 2003).

The concept of individual differences in rhythm perception strategies has been investigated by McAuley and colleagues. In a study of 43 students individual differences were found using a rhythm judgment task (detailed below). Participants showed different response patterns in reaction to the same stimuli (McAuley et al., 2006). A functional magnetic resonance study reported correlations between responses in the rhythm judgment tasks and different patterns of activity in the auditory and motor areas of the cerebral cortex (Grahn & McAuley, 2009). Another fMRI study using the same task reported a connection between decreased beat induction (as a characteristic feature of entrainment) and the neural activity in the basal ganglia, premotor and supplementary motor regions, and thalamus.

In the procedure developed by McAuley et al. (2006), participants listened to short tone sequences and judged whether the sequences were 'speeding up' or 'slowing down' in the end. The sequences consisted of four or five 50ms piano tones with 440Hz fundamental pitch. The timing of tones in the sequences is illustrated in Figure 1.

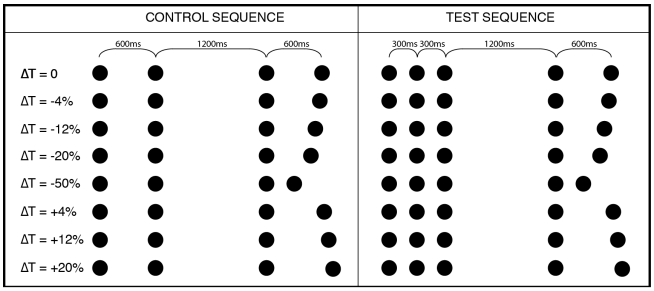


Figure 1. Layout of tones in control and test sequences in the rhythm perception task developed by McAuley et al. (2006).

Control (four tone) sequences consisted of two tones 600ms apart, followed by a 1200ms pause, followed by two tones with variable time interval (600ms ± ΔT). Test (five tone) sequences consisted of three tones 300ms apart, followed by 1200ms pause, followed by two tones with the same variable time interval (600ms ± ΔT). The final tone varied in eight possible intervals: T (600ms), T/2 (300ms), T±4% (576ms or 624ms), T±12% (528ms or 672ms) and T±20% (480ms or 720ms). Overall, the stimuli consisted of sixteen possible sequences, eight for control and eight for test conditions.

Depending on the rhythm perception strategy, participants should evaluate the test sequences differently. Weak beat perceivers (associated with the interval-based strategy) should perceive those sequences as slowing down, regardless of ΔT, as the inter-onset intervals (IOIs) between first three tones are shorter than consecutive IOIs. On the other hand, strong beat perceivers (associated with entrainment strategy) should evaluate test sequences according to ΔT (speeding up when $\Delta T < 0$ and slowing down if $\Delta T > 0$). This is due to hypothetical induction of the sense of 600ms beat when using this strategy. Control sequences should yield no significant differences, and participants are expected to assess those sequences similarly, regardless of their rhythm perception strategy.

4. PRESENT STUDY

This study investigates possible sources of individual differences in rhythm perception strategies. While previous research showed that participants had different rhythm perception strategies (McAuley et al., 2006) and these strategies correlated with different patterns of brain activity (Grahn & McAuley, 2009; McAuley, Henry, & Tkach, 2012), there is no clear evidence as to where those differences stem from.

In general, individual differences may originate from innate, genetic factors, or previous experience. This study focuses on the possibility that individual differences in rhythm perception strategies are associated with musical training. Any type of musicianship, whether it is singing or playing an instrument, requires relatively precise timing. The timing of played notes in most cases needs to be adjusted to a specific tempo (provided by a conductor, drummer or a metronome). This ability to adjust to a given tempo or rhythm can be in fact described as the process of synchronization, or *entrainment*, to an external oscillation. Thus, we hypothesize that subjects with previous musical training should prefer rhythm perception strategies based on entrainment more than subjects without prior musical experience. To test the aforementioned hypothesis, we performed a rhythm perception task developed by McAuley et al. (2006) in two groups – a group of music students with more than 10 years of formal musical training, and a control group of university students without any experience in playing music.

5. PARTICIPANTS

A group of twenty-two students of Stanisław Moniuszko Academy of Music in Gdańsk, Poland was recruited for the study. The group consisted of five women and seventeen men. Age varied between 19 and 35 (M=24.13; SD=4.36). All participants had more than 10 years of formal musical training (M=12.5; SD=1.60). A control group of twenty-five students with no musical training was recruited from students of the University of Gdańsk and Higher School of Security in Gdańsk. In this group, age varied between 19 and 49 (M=26.53; SD=7.03).

6. METHODS

The stimuli were presented to participants using a computer running a custom-developed application written in Java (Oracle Corporation). The stimuli were presented binaurally using Sony Professional MDR-7510 headphones (Sony Corporation) at a comfortable listening level (participants set the level before the task). The entire procedure consisted of a familiarization block of sixteen trials (one for each possible sequence, no results were gathered), followed by a test block of forty-eight trials (three for each possible sequence). The trials were randomized independently for each subject. Participants were asked to judge whether the sequences they heard were speeding up or slowing down towards the end and gave their responses using a touchpad and two buttons that appeared on the screen.

7. RESULTS

The distributions of probabilities of 'speeding up' responses in control sequences for two groups are shown in Figure 2. These probabilities were computed following the methodology proposed by Grahn & McAuley (2009), based on signal detection theory (Macmillan & Creelman, 1991). Predicted proportions of 'speeding up' responses for each sequence were generated using cumulative normal distribution function, $P = 1 - \Phi(z)$.

Values in the vertical axis represent the probability of 'speeding up' response when closer to 1 and 'slowing down' when closer to 0. Values near 0.5 indicate equal probability of 'speeding up' and 'slowing down' (dotted line). Results of control sequences show a generally predictable trend, with probability of 'speeding up' responses higher in shorter final IOIs, and 'slowing down' responses higher in longer final IOIs. Interestingly, most respondents judged the 600ms ($\Delta T = 0$) interval as slowing down, regardless of prior musical training.

Control sequences yielded significant differences between groups in 480ms ($\Delta T = -20\%$), 528ms ($\Delta T = -12\%$), 576ms ($\Delta T = -4\%$) and 720ms ($\Delta T = +20\%$). In all but one cases, musicians heard the sequences as 'speeding up' more often than non-musicians. Only in the $\Delta T = +20\%$ condition, musicians heard it as 'slowing down' almost universally, while some non-musicians assessed it as 'speeding up'. The results of the t-test for differences between mean responses of musicians and non-musicians in control sequences are presented in Table 1.

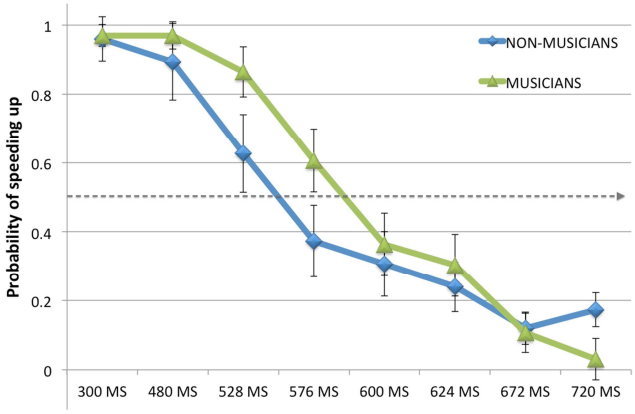


Figure 2. Probabilities of 'speeding up' responses in control sequences for musicians and non-musicians.

Table 1. Results of the t-test for differences between mean responses of musicians and non-musicians in control sequences.

	Non-musicians		Musicians		t-Test	
	M	SD	M	SD	t	p
Ctrl. 300ms	.96	.11	.97	.10	-0.316	.377
Ctrl. 480ms	.89	.16	.97	.10	-1.951	.029
Ctrl. 528ms	.63	.28	.86	.20	-3.333	.002
Ctrl. 576ms	.37	.28	.37	.34	-2.602	.013
Ctrl. 600ms	.31	.32	.61	.29	-0.639	.526
Ctrl. 624ms	.24	.28	.36	.32	-0.715	.478
Ctrl. 672ms	.12	.19	.11	.19	0.252	.802
Ctrl. 720ms	.17	.26	.03	.10	2.457	.015

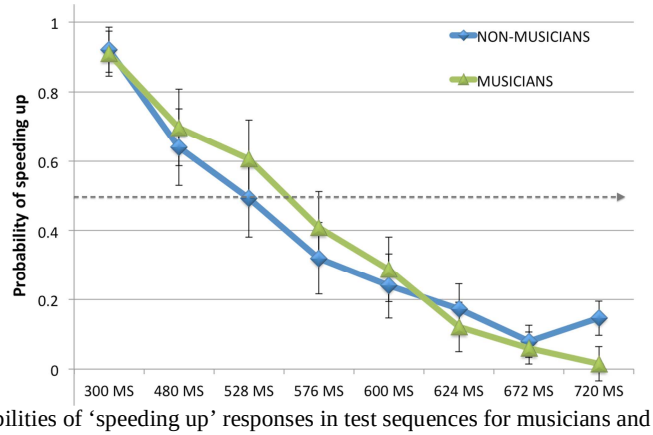


Figure 3. Probabilities of 'speeding up' responses in test sequences for musicians and non-musicians.

Table 2. Results of the t-test for differences between mean responses of musicians and non-musicians in test sequences.

	Non-musicians		Musicians		t-Test	
	M	SD	M	SD	t	p
Test 300ms	.92	.17	.91	.26	,173	,864
Test 480ms	.64	.35	60	.40	-,526	,602
Test 528ms	.49	.39	.61	.37	-1,023	,312
Test 576ms	.32	.33	.41	.37	-,878	,385
Test 600ms	.24	.31	.29	.31	-,524	,603
Test 624ms	.17	.26	.12	.22	,743	,461
Test 672ms	.08	.17	.06	.13	,426	,672
Test 720ms	.15	.22	.02	.07	2,716	,008

The distribution of probabilities of responses in the test sequences is shown in Figure 3. Compared to control sequences, the responses in test sequences for both groups are closer to the 0.5 level, which indicates less certainty in the assessment of those sequences. No significant differences between groups were found in the test sequences except for $\Delta T = +20\%$ condition. This result is contrary to our expectations, as differences in rhythm perception strategies between musicians and non-musicians were expected to emerge in test sequences. A significant difference in the $\Delta T = +20\%$ condition can be explained by almost universal certainty of the stimulus slowing down in the musician group. The results of the t-test for differences between mean responses of musicians and non-musicians in test sequences are presented in Table 2.

8. DISCUSSION

The results of this study suggest, somewhat surprisingly, that rhythm perception strategies are not related to musical training. While the experiment revealed significant differences between musicians and non-musicians in one test condition (720ms, $\Delta T = +20\%$), mean values in this condition suggest that both musicians and non-musicians perceive it generally as slowing down (only with greater certainty in the musician group). Significant differences in the control sequences seem to be determined by greater certainty of given responses among musicians. Non-musician group is mostly characterized by a greater ambiguity of answers (results closer to 0.5 line).

Overall, the lack of differences between groups in test sequences coupled with relatively high variance of responses suggest, that participants do use different rhythm perception strategies. This is however independent of previous formal musical training.

If individual differences rhythm perception strategies are not determined by previous musical training than what are they determined by? This question remains open for further enquiry. One possible explanation is that they are determined biologically by innate capabilities of human nervous system. This hypothesis is sanctioned in a way by research on the spontaneous motor tempo (SMT). There is some evidence, that SMT is genetically conditioned, as identical twins have been reported to have more similar SMT compared to fraternal twins (Fraisse, 1982). This innate predisposition may influence higher levels of cognitive analysis of rhythm, such as perception strategies.

Another possible explanation of individual differences in rhythm perception is a more situational one. The same subjects may use different perceptual strategies depending on current context, mood, or even a specific piece of music they have listened to just before the experiment. To verify this hypothesis, research on the stability of rhythm perception strategies over time should be conducted. If individual differences in perception strategies were proved to be unstable over time, it would point to new research directions concerning situational factors that influence perceiving rhythm.

9. CONCLUSIONS

The results of this research provided evidence that individual differences in rhythm perception strategies are not determined by previous musical training. Future research in this area is recommended, in order to explore other potential explanations for the existence of individual differences in perceiving rhythm.

10. REFERENCES

Chen, A. (1983). Tones of the oscillating sun. Science News, *123(25), 392 – 395.*

Fernald, A., & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J. L. Morgan & K. Demuth (Eds.), From signal to syntax: Bootstrapping from speech to grammar in early acquisition *(pp 365-388). Hillsdale, NJ: Erlbaum.*

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. Cognition, 100*(1), 173–215.*

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), The Psychology of Music *(pp. 149–180). Orlando, FL: Academic.*

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. Psychological Review, 84*(3), 279–325.*

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., Scott, S.K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. Proc Natl Acad Sci U S A, *99(16), 10911-6.*

Grahn, J. A. (2012). Neural mechanisms of rhythm perception: current findings and future perspectives. Topics in Cognitive Science, 4*(4), 585–606.*

Grahn, J. A. & McAuley, J. (2009). Neural bases of individual differences in beat perception. NeuroImage, 47*(4), 1894–903.*

Harrer, G. (1975). Grundlagen der Musiktherapie und Musikpsychologie. *Stuttgart: Gustav Fischer Verlag.*

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition, *92(1–2), 67–99.*

Honing, H., & Ploeger, A. (2012). Cognition and the Evolution of Music: Pitfalls and Prospects. Topics in Cognitive Science, *4 (4), 513-524.*

Keele, S. W., Nicoletti, R., Ivry, R. I., & Pokorny, R. A. (1989). Mechanisms of perceptual timing: Beat-based or interval-based judgements? Psychological Research, 50*(4), 251–256.*

Large, E., & Jones, M. (1999). The dynamics of attending: how people track time-varying events. Psychological Review, 106*(1), 119–159.*

Macmillan, N.A., Creelman, C.D., 1991. Detection Theory: A User's Guide. *Cambridge University Press: New York.*

McAuley, J. (2010). Tempo and rhythm. In M. Jones, R. R. Fay, & A. N. Popper (Eds.), Music Perception. Springer Handbook of Auditory Research, 36, *165–199. Springer Science + Business Media.*

McAuley, J., Frater, D., Janke, K., & Miller, N. S. (2006). Detecting changes in timing: Evidence for two modes of listening. The Proceedings of the 9th International Conference on Music Perception and Cognition, *566-573.*

McAuley, J., Henry, M. J., & Tkach, J. (2012). Tempo mediates the involvement of motor areas in beat perception. Annals of the New York Academy of Sciences, 1252, *77–84.*

McAuley, J., & Jones, M. R. (2003). Modeling Effects of Rhythmic Context on Perceived Duration: A Comparison of Interval and Entrainment Approaches to Short-Interval Timing. Journal of Experimental Psychology. Human Perception & Performance, 29*(6), 1102–1125.*

Overy, K. (2000). Dyslexia, Temporal Processing and Music; The potential of music as an early learning aid for dyslexic children. Psychology of Music, *28(2), 218-229.*

Patel, A. D. (2007). Music, Language, and the Brain. *Oxford: Oxford University Press.*

Patel, A. D., Iversen, J. R., Bregman, M. R., Schulz, I., & Schulz, C. (2008). Investigating the human-specificity of synchronization to music. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), Proceedings of the 10th Intl. Conf. on Music Perception and Cognition, August 2008, Sapporo, Japan. *Adelaide: Causal Productions.*

Repp, B. (2007). Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. Cognition, *102(3), 434–454.*

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. Journal of Memory and Language, *49, 249–267.*

[2-23] AMBIGUITY OF TEMPO PERCEPTION FOR J-POP AND NON J-POP TUNES

Sota Okada[1], Hiroshi Kuwabara[2], Masanobu Miura[1]
[1]Faculty of Science and Technology, Ryukoku University, Japan, [2]Graduate School of Science and Technology, Ryukoku University, Japan
Poster

ABSTRACT

This study investigates the ambiguity of tempo perception in different musical genres. Ten students listened to 100 tunes of J-POP (Japanese POPular) and 100 tunes of non J-POP and rated the tempo value for each tune. The tunes were then categorized into clusters by using the *k*-means algorithm for tempo values that the ten participants rated. Then the authors investigated the commonality of rated tempo values. That is, the authors compared the ambiguity of the tempo perception for J-POP and non J-POP. In the results, the number of tunes for which ten participants perceived an almost common tempo was 50 out of 100 J-POP tunes but only 11 out of 100 non J-POP tunes. Therefore, this confirms that non J-POP has more ambiguity of tempo perception than J-POP.

1. INTRODUCTION

The field of Music Information Retrieval (MIR) has been developed to retrieve music information such as tempo or beat automatically [1-5] from musical audio. Specifically, many studies have focused on tempo estimation. They estimated one tempo value for a tune [1-5]. In contrast, recent studies on tempo estimation have concentrated on the perceptual variety of tempo. Studies for tempo perception have reported the ambiguity of tempo perception among people [6-7]. The ambiguity of the tempo perception is, however, not widely understood in MIR, and studies that take it into consideration are rarely conducted.

This study investigates the ambiguity of the tempo perception in different musical genres: J-POP and non J-POP. Section 2 describes problems in previous studies. Section 3 explains a scheme to investigate the ambiguity of the tempo perception in different musical genres. Section 4 details an experimental investigation where participants rated tempo values for J-POP and non J-POP tunes. Furthermore, in section 5, we establish a standard to determine the extent of the ambiguity of the tempo perception to investigate the variety of perceived tempo for J-POP and non J-POP tunes. Then, we classify the extent of the ambiguity of tempo perception on the basis of our original standard. Finally, we discuss the results in section 6 and conclude the paper in section 7.

2. PROBLEMS IN RELEVANT STUDIES

The relevant studies of the ambiguity of the tempo perception [6-7] predicted agreement and disagreement in the tempo perception [6] and investigated the range of tempo values that is likely to be perceived [7]. Peeters *et al.* [6] proposed an algorithm to predict the commonality of the tempo perception on the basis of physical features of the acoustic waveform. McKinney *et al.* [7] reported that the tempo values most likely to be perceived ranged of 87 [bpm] to 175 [bpm] and that nearly 120 [bpm] is the most perceived of all tempo values. These studies, however, covered only the ambiguity of the tempo perception in different tunes. They investigated neither the differences in genres nor factors of the tempo perception ambiguity. In this study, we focus on musical genres as a factor of the ambiguity and investigate the ambiguity of the tempo perception in them.

3. AMBIGUITY OF TEMPO PERCEPTION

3.1. Investigation scheme of ambiguity of tempo perception

Other than certain specific genres such as ethnic music, people listen to tunes in which beats are possible to perceive. Therefore, it is possible to perceive the tempo values from the cycle of beats. The tempo perception differs for each person on the basis of the double or triple from the cycle of beats. Even though we investigate the ambiguity of the tempo perception in different genres using a listening experiment, whether the ambiguity of the tempo perception is derived from either a genre difference or a personal difference is still under discussion. In other words, to investigate the ambiguity of the tempo perception by the difference in genres, we need to compare the varieties of perceived tempo values among genres, where the variety of perceived tempo values in the reference genre is hopefully consistent among people. Therefore, we focused on J-POP. Since J-POP has become more popular internationally in recently years, participants are not expected to perceive a tempo very differently. Moreover, J-POP tunes are mostly made for commercial purposes. Using J-POP tunes may guarantee that there is little personal difference in the tempo perception. Thus, this study employs a strategy that compares J-POP with non J-POP.

3.2. Summary of investigation

This study investigates the difference in the ambiguity of the tempo perception by conducting a listening experiment using J-POP and non J-POP tunes. Specifically, a listening experiment for the tempo perception was conducted by ten participants, who listened to tunes of J-POP and non J-POP and were asked to rate their tempo values. Then, we calculated the number of people who perceived a common tempo for the same tune.

The following is a tentative standard, but these sorts of criteria have not been reported up to now: when eight to ten people perceive a common tempo, the tune is classified as "common tempo tune (CT)". When five to seven people do so, the tune is classified as "partial common tempo tune (PCT)". When zero to four people do so, the tune is classified as "non-common tempo tune (NCT)". Also, we regard the PCT and the NCT as those tunes that have tempo perception ambiguity. We investigate the ambiguity of the tempo perception to classify J-POP and non J-POP tunes into each catagory.

4. LISTENING EXPERIMENT

4.1. Experimental Conditions

4.1.1. Participants

We conducted a listening experiment using ten students of our university: nine males and one female. Their average age was 21.4 (SD=0.7). All participants' were native Japanese speakers.

4.1.2. Tunes

The listening experiment used 100 J-POP tunes and 100 non J-POP tunes. Non J-POP were defined as those sung by non-Japanese language singers.

4.2. Method

Participants rated perceived tempi three times for each tune. They freely play back from arbitrary position in a tune. Tunes were played on a desktop computer (FMV ESPRIMO, Fujitsu Corp) and listened to using headphones (MDR-CD900ST, Sony Corp). The sound pressure level when listening to tunes was determined by each participant for his/her comfort. Participants were allowed to use an electronic metronome (Digital Tuner Metronome, KORG Corp) that calculates tempo by observing the intervals of tapping. Participants rated tempi three times for a tune, and we extracted the two closest tempi of the three. We then calculated the average of the two tempi. We defined the average of the two tempi as the perceived tempo value by the participant for the tune. Figure 1 shows the flowchart for calculating the perceived tempo value for a tune.

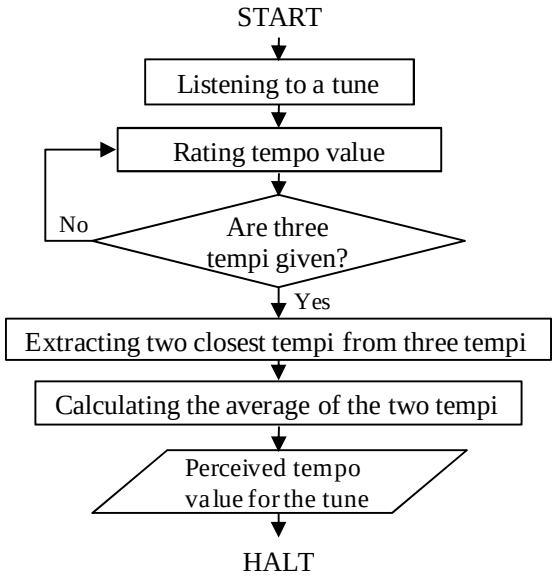


Figure 1: Flowchart for calculating the perceived tempo value for a tune.

5. METHOD OF CLASSIFICATION

5.1. Method to determine common tempo tune

We obtained ten tempo values for each tune. Then, we classified the tunes into CT, PCT, or NCT. To do this, we established a standard, since no objective index to determine whether an obtained tempo value is common or not has yet been reported. Here the commonality of perceived tempo values is evaluated in accordance with the standard. Specifically, we calculate the average of the ten tempo values provided by the ten participants. When ten tempo values are within the range of plus minus 8% of the mean of the ten, the tune was classified as CT.

5.2. The *k*-means algorithm

If any of the ten tempi is out of the range of plus minus 8%, a method must be provided to determine whether it is PCT or NCT. To do so, we propose a method based on the *k*-means algorithm.

The *k*-means algorithm is a non-hierarchical clustering algorithm and is widely used as an unsupervised clustering algorithm. In the first step of the *k*-means algorithm, each sample is randomly and tentatively categorized into a *k* cluster. Then, the algorithm calculates the centroid of each cluster and again classifies each sample into the nearest cluster by observing the nearest centroid. Furthermore, the algorithm keeps doing this step until no sample moves to another cluster. In short, the *k*-means algorithm determines to which cluster each sample belongs.

5.3. Method for classifying tunes by using the *k*-means algorithm

The method in section 5.1 extracts all CT tunes, so we classify the rest into CT, PCT, or NCT by using the *k*-means algorithm. We first calculate the logarithm of ten tempi perceived by ten participants. We assume the case where any ten tempi perceived by ten participants are not included within the range of plus/minus 8% of the calculated mean. We try to solve this problem by the *k*-means algorithm, but another problem concerning the number of cluster *k* may happen. For example, when one participant perceives 60 [bpm], eight perceive 120 [bpm], and the other perceives 100 [bpm], we try to classify the tempi into two clusters (*k*=2). Thus, the eight who perceive 120 [bpm] and the one who perceives 100 [bpm] are classified into the same cluster, which is not appropriate since the ratio 120/100 is larger than 8%. Therefore, if we classify each tempo by using the *k*-means algorithm, the *k* should be changed into another value, and we should be able to confirm whether each tempo was correctly classified into clusters or not. When we investigated how ten participants perceived the tempo values for 100 J-POP and 100 non J-POP, we confirmed that each tempo is classified into four clusters at most. Therefore in this study, the *k* is changed from 2 to 4. Then, we calculate the classification results of three patterns of two (*k*=2), three (*k*=3), or four (*k*=4). In the case of $2 \leq k \leq 4$, we calculate the average tempo values as y_{jk} in *j*th cluster and calculate the average difference among clusters as $\bar{y}_k$.When $\bar{y}_k$, has a maximum value among $\bar{y}_2$, $\bar{y}_3$, and $\bar{y}_4$, the k' is employed as to be the number of appropriate clusters. Eq. (1) shows y_{jk}.

$$y_{jk} = \frac{\sum_{i=1}^{N_{jk}} \log t_i}{N_{jk}} \quad (1)$$

In eq. (1), k (k=2,3,4) is the number of clusters, j (j=1,2,…,k) is a cluster number, N_{jk} is the number of samples in jth cluster, and t_i (i =1,2,…,N_{jk}) is the tempo value classified into each cluster. Eq. (2) shows $\bar{y}_k$.

$$\bar{y}_k = 2\frac{\sum_{j=1}^{k-1}\sum_{j'=j+1}^{k}\left|y_{jk} - y_{j'k}\right|}{k(k-1)} \quad (2)$$

We extract the cluster that has the maximum number of samples from clusters and count up the number of tempo values classified into the cluster. Then, we calculate the number of people who perceived similar tempi for tunes and classify the tune into CT, PCT, or NCT. Figure 2 shows the flowchart for classifying tunes by the perceived tempo value.

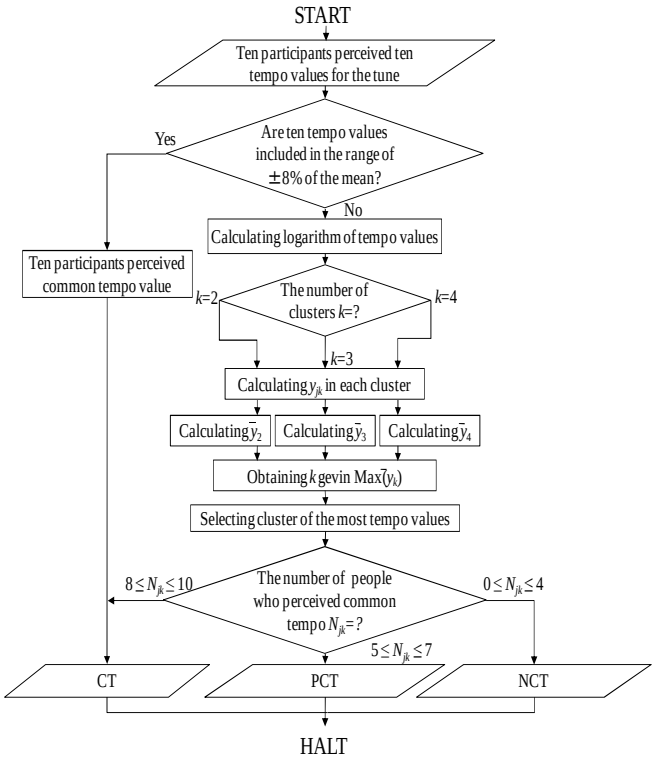


Figure 2: Flowchart for classifying tune by the perceived tempo value

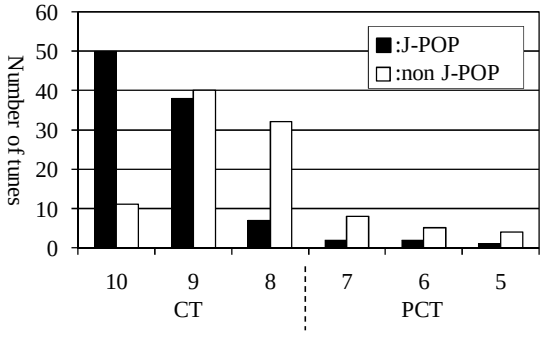


Figure 3: Results of the number of people that perceived a common tempo

6. RESULTS AND DISCUSSION

6.1. Results

We classified tunes on the basis of the method described in section 5. Figure 3 shows the number of people who perceived a similar tempo. In figure 3, ten participants perceived a common tempo value for 50 out of 100 J-POP tunes but only for 11 out of 100 non J-POP tunes. We confirmed that at least five people perceived a common tempo value for both J-POP and non J-POP. In other words, NCT was not observed for both J-POP and non J-POP, and we confirmed that the number of PCT for J-POP is less than non J-POP (five J-POP tunes and 19 non J-POP tunes). Therefore, we confirmed that non J-POP has more ambiguity of tempo perception than J-POP.

6.2. Discussion

In section 6.1, we confirmed that J-POP has less ambiguity of tempo perception than non J-POP. A possible factor in this result is the familiarity of tunes. All ten participants in the listening experiment were native Japanese speakers who probably have more opportunities to listen to J-POP than non J-POP since people can be exposed to J-POP regularly in Japan. Another factor is the lyrics of the tune, since the participants were likely to listen to and grasp the meaning of lyrics while listening and establish the tempo on the basis of the speed of lyrics. These are just our speculations, but we will discuss these phenomena more. In this study, we definitely confirmed the existence of ambiguity of tempo perception difference in different musical genres.

7. CONCLUSION

This study investigated the ambiguity of tempo perception for J-POP and non J-POP. Specifically, a listening experiment for tempo perception was conducted by ten participants, who listened to 100 J-POP tunes and 100 non J-POP tunes. We calculated the number of tunes for which participants perceived a common tempo value and investigated the ambiguity of the tempo perception between both genres. In the results, we confirmed that non J-POP has more ambiguity of tempo perception than J-POP. For future works, we will consider the ambiguity of the tempo perception by using another genre and physical parameters that relate to the ambiguity of the tempo perception.

8. ACKNOWLEDGMENTS

This study is partly supported by MEXT, Kakenhi (25580050).

9. REFERENCES

Xiao, L et al., *"Using a statistic model to capture the association between timbre and perceived tempo," Proc. ISMIR, 2008.*

Klapuri, A et al., *"Analysis of the meter of acoustic musical signals," IEEE Trans. Speech and Audio Processing, 14(1), 2006.*

Scheirer, E., "Tempo and beat analysis of… acoustic musical signals," J. Acoust. Soc. Am., 103(1), 1998.

Dixon, S., "Automatic extraction of tempo and beat from expressive performances," Journal of New Music Research, 30 (1), 2001.

Alonso, M et al., *"Tempo and beat estimation of musical signals," Proc. International Conference on Music Information Retrieval, Barcelona, 2004.*

Peeters, G et al., *"Predicting agreement and disagreement in the perception of tempo," Proc. International Symposium on Computer Music Multidisciplinary Research, pp.253-266, 2013.*

McKinney, M.F et al., *"Ambiguity in tempo perception: what draws listeners to different metrical levels?", Music Perception: An Interdisciplinary Journal, Vol.24, No.2, pp.155-166, 2006.*

[2-24] AN INQUIRY INTO HEARING-IMPAIRED STUDENT'S MUSICAL ACTIVITIES − HOW DO THEY LISTEN TO THE MUSIC?

Masaki Matsubara[1], Hiroko Terasawa[1], Kjetil F. Hansen[2], Rumi Hiraga[3]
[1]University of Tsukuba, Japan, [2]Kth Royal Institute of Technology, Sweden, [3]Tsukuba University of Technology, Japan

Poster

1. Background: Hearing impairments make listening to sounds harder in many respects, but musical activities still have positive effects and offer positive experiences for persons with hearing disabilities. Several studies have reported about musical activities of persons with more than 100dB hearing-loss, and implied the effectiveness of music for hearing impaired in terms of music therapy, education and language acquisition [Darrow2006, Torppa2010, Hansen2011]. However, it is still uncertain how specific musical features and structure are appreciated and recognized in music listening.

2. Aims: By understanding and improving the conditions for listening for hearing-impaired persons, quality-of-life will also improve. We have already developed a music game for training perception by active listening [Hansen2013, Hiraga2013]; to validate the appropriateness of the game we need to examine the details of how hearing-impaired persons listen to music, including recognizing the activities and situations involved in listening.

3. Method: We conducted a questionnaire study to investigate hearing-impaired students' listening habits. Twenty persons (11 males, 9 females) with congenital hearing disabilities volunteered to participate and answer to questions about their musical activities. The questionnaire consisted of questions about their listening preferences, listening environment, what kind of activities they engage in, musical experience, and about situations in which music listening is experienced as being difficult.

4. Results: According to the responses, only twelve participants liked listening to music, six were neutral, and two disliked music listening. Most assumed that "music" was vocal-based songs, although all had experience of instrumental music from elementary school or junior high school. Interestingly, the participants mentioned music containing high-pitched singing or choir singing to be the hardest to listen to.

5. Conclusions: Most of the participants have musical experience and like listening to music. Due to the hearing impairment, the experience of high frequency sounds is limited, and possibly as a consequence of this, high-pitched singing was found to be difficult. Training the listening abilities can be helpful in this regard. To accommodate training we need to investigate further how high-frequency sounds in music is perceived, and how hearing-impaired persons recognize musical structure (e.g. melody, rhythm, song).

[2-25] EFFECTS OF PAIR ENSEMBLE ON ATTENTION AND AFFECTIVE EXPERIENCE

Hitoshi Tominaga[1], Teruo Yamasaki[2], Yukiko Uchida[3], Yuri Miyamoto[4]
[1]Graduate School of Human & Environmental Studies, Kyoto University, Japan, [2]Faculty of Psychology, Osaka Shoin Women's University, Japan, [3]Kokoro Research Center, Kyoto University, Japan, [4]Department of Psychology, University of Wisconsin - Madison, USA
Poster

In this study, we explored what effects ensemble performance may have on attention and on affective experience. Previous studies suggested that the cultural differences exist in contextual information sensitivity. Individuals in Asian cultural contexts are more likely to incorporate contextual information, whereas those in North American cultures are more likely to ignore contextual information (e.g., Masuda & Nisbett, 2001). We hypothesized that "collective achievement" in cooperative task encourages the capacity of incorporating contextual information in order to adjust own performance to other's performance. In a cooperative situation, people who failed the collective task attended more holistically in order to maintain cooperative relationships. Moreover, for social oriented individuals, cooperation task may raise their self-esteem.In our experiment, participants were asked to conduct ensemble tasks. Participants played the xylophone with a partner (pair-ensemble condition: the achievement seems to be collective) or with PC (PC-ensemble condition: the achievement seems to be personal). On each condition, we devided the participants into two groups (success vs. failure) according to their self-evaluation of the ensemble task. In addition, we computed objective achievement from recorded data. Then we conducted the Framed Line Test (Kitayama, Duffy, Kawamura & Larsen, 2003) to measure contextual information sensitivity, and asked to answer questionnaires about social orientedness. Furthermore, we collected the data about self-esteem both before and after the ensemble tasks, and calculated raise of self-esteem during conducting the ensemble tasks. As a result, we found a significantly larger error difference between absolute tasks and relative tasks in the collective failure group than in the collective success group. Therefore, participants in the collective failure group attended more holistically than participants in the collective success group. Additionally, in the pair-ensemble condition, social orientedness was positively correlated with raise of self-esteem and objective achievement. Therfore, social oriented participants were good at performing pair ensemble, and enhanced own self-esteem through the ensemble performance. These results provide indirect evidence regarding the importance of collectiveness and sociality for ensemble performance.

[2-26] THE INFLUENCE OF TASK AND CONTEXT ON COMPLEX RHYTHM PRODUCTION: EVIDENCE FROM MALIAN DRUMMING.

Justin London[1], Rainer Polak[2]
[1]Carleton College, USA, [2]Hochschule für Musik und Tanz, Köln, Germany
Poster

1. Background: It has long been presumed that simple integer ratio rhythms ("SIRRs," e.g., 1:1, 2:1) are easier to produce, discriminate and synchronize with than more complex, non-integer ratio rhythms ("NIRRs"—Fraisse 1956, Povel 1981, Summers et al 1989). Yet Desain and Honing (2003) found that centroids of perceptual categories were not SIRRs, and Repp, London, & Keller (2011, 2012) showed NIRRs are produced as accurately as SIRRs, though both were subject to slight distortions in production, as they found an attractor ratio for 2-element NIRRs of ≈1.32:1.

2. Aims & Method: To examine NIRRs in a real-world setting, timing data from Malian drumming performances representing different genres, players, and contexts were collected. All involve NIRRs as a basic metrical framework and all undergo a significant global accelerando. From this data the stability and coordination of NIRRs across pieces, performers, performances, and tempos can thus be studied.

3. Main contribution: NIRRs found both two- and three-element Malian rhythms ranging between 1.57:1 to 1.32:1, congruent with Repp, London, & Keller (2011). These NIRRs are stable (SD of variance ≤ 3%) even under significant tempo changes, contra Repp, London, & Keller (2005) and Snyder, et al (2005) who found NIRRs tending to devolve to SIRRs in the absence of a pacing signal. NIRRs produced by Malian drummers include complex elaborations of basic patterns, which did not affect their primary ratios, contra Repp, London, & Keller (2012, experiment 3). Onsets between parts were closely aligned (≤ 10ms asynchrony).

4. Implications: The NIRRs produced by Malian drummers fall between 1:1 and 2:1, a response to the musical task "play a rhythm that is neither a duplet or triplet." This data also shows that SIRRs do not give an advantage in synchronization, supporting Himberg (2014) who found that coordination between live performers is often better than with a mechanical pacer. This data also reinforce the importance of studying rhythm production in ecologically valid contexts in order to gain a true picture of human rhythmic capabilities.

[2-27] A CLINICAL STUDY ABOUT THE EFFECT OF MUSIC THERAPEUTIC TREATMENTS IN AN ONCOLOGICAL REHABILITATION

Eun-Jeong Lee[1], Jens-Peter Rose[1], Joachim Weis[1]
[1]Tumor Biology Center Freiburg, Germany
Poster

1. Background: In the context of psycho-oncology, music therapy is one of various complementary treatments used to support cancer patients psychologically as well as physically. This study investigated the effect of music therapeutic treatments on cancer patients during their period of rehabilitation, focussing in particular on any positive change in body sensitivities, mental state, and emotions.

2. Method: This clinical study has been conducted with 126 cancer patients who received oncological rehabilitation treatments over the course of 3-4 weeks. 3 different types of music therapeutic treatments were performed in different groups: sound meditation group (n=40), improvisation group (n=43) and drumming group (n=43).

3. Results: After each music therapeutic session, the patients' body sensitivities and mental state had significantly improved. Further, the expressivity and perception of emotion were partially changed. Conclusion: This clinical study shows that music therapy can be an important part of the psycho-oncological concept for oncological rehabilitation.

Poster Session 3

Baekyang Building Lobby
16:15 – 17:00, Thursday 7 Aug 2014

[3-01] EMOTIONAL REASONS FOR PERSONAL MUSIC LISTENING: FREQUENCIES AND AFFECTIVE OUTCOMES

William Randall[1], Nikki Rickard[1]
[1]Monash University, Australia

Poster

An important aspect of researching everyday music use is understanding the reasons people have for listening to music. While this has been the focus of an extensive body of research, findings have been inconsistent, and the frequencies and affective outcomes of different reasons for listening remain unclear. Emotional reasons for listening are of particular interest, as these have been consistently shown to be of central importance to everyday music use. The current study aimed to provide empirical data to clarify the frequencies of reasons for listening, and their affective outcomes, by using the experience sampling method (ESM). 327 participants used the MuPsych app, a mobile ESM designed for the real-time and ecologically valid measurement of personal music listening. Results revealed that emotional reasons for listening were used more frequently in a negative affective state, and often produced undesirable affective outcomes. In particular, listening to cope with a situation or to forget problems was associated with poor emotional health and well-being, and failed to improve any affective state. It was concluded that while personal music listening is generally beneficial, the initial mood of the listener, combined with certain emotional reasons for listening, can lead to undesirable outcomes.

[3-02] LISTENER RESPIRATION: ACTIVE, ATTENTIVE, OR DISTRACTED

Finn Upham[1]
[1]New York University, USA

Poster

1. Background: What kind of respiratory behaviour arises while subjects listen to music? Subjects breathing reflects their actions and imagined actions, with greater regularity during attentive listening to speech (Brown, 1962). Parameters such as inhalation/exhalation ratio mark salient differences between speech tasks with different cognitive demands, even when subjects only imagine speaking (Conrad & Schönle, 1979). Since subjects show measurable consistency in when they breath upon repeated listening to some musical stimuli (Sato, Ohsuga, and Moriya, 2012; Upham, 2013), what disruptions of respiratory periodicity result in such alignment?

2. Aims: This paper presents criteria for differentiating categories of listening behaviour to recognise periods of high regularity (associated with attentive listening), short disruptions (yawns, sneezes, sighs), and high irregularity (associated with imagined vocalisation, dancing, and crying), and considers which conditions support respiratory phase alignment with musical stimuli.

3. Method: In thirty hours of respirating band data from a single subject (the author)while listening to musical stimuli), each breath is described in terms of duration, inhalation/exhalation ratio, relative depth, phase distortion from median breath contour, and tagged with respiration behaviour according to annotations from experimental sessions. Possible indications of attentive listening are evaluated per listening, from the distributions of the breath parameters. From previous work on coordination of respiratory phase across repeated listenings on this same data set, moments preceding high alignments are considered in terms of phase distortion and respiration category to suggest mechanisms supporting intermittent stimulus-respiratory phase alignment to musical stimuli.

4. Results: Initial results show that the distribution of inhalation/exhalation ratios depend markedly with respect to the stimulus, consistent with the idea that different stimuli provoke different listener behaviour, including irregular respiratory behaviour indicative of different forms of active engagement. In both disruptive and regular periods of respiratory behaviour, phase adjustments are often made by shortening or lengthening the stable interval preceding the onset of the next inhalation.

5. Conclusions: Respiratory regularity of music listeners may be useful indicators of listening behaviour, whether attentive listening or active expressive behaviour (affective or musical). Since respiration can be tracked passively, this may be a useful means of assessing listener engagement and listening strategies without self report.

[3-03] AUTOMATIC ESTIMATION OF RELEASED DECADES OF JAPANESE-ANIME MUSIC FROM MUSICAL AUDIO

Mio Sakurai[1], Masanobu Miura[2]
[1]Graduate School of Science and Technology, Ryukoku University, Japan, [2]Faculty of Science and Technology, Ryukoku University, Japan

Poster

ABSTRACT

This paper describes an automatic estimation of decades Japanese-anime music (JAM) was released from musical audio. Relationships between music and ages were previously investigated, and a problem was found that music covering several genres may affect the clarified results. Therefore, the present study focuses on a specific musical genre, JAM, and investigates the change in it over decades on the basis of the musical audio. The musical audio of the JAM are firstly collected. Next 25 types of MIR parameters are calculated for the collected musical audio. Then an automatic recognition of decades was carried out, where the 25 parameters are used in machine recognition, and the accuracy rate and effectiveness of estimation were obtained. The accuracy rates of each parameter in single or double decades are then estimated. Released decades were then subjectively evaluated by ten subjects who were asked to guess the decade released for 60 tunes of JAM. Finally, this paper discusses the correspondence relationships between released decades of JAM and MIR parameters of musical audio on the basis of the results. As a result, the authors confirm that released decades are automatically estimated using the squeezed parameters with high precision and that the method performs comparably to people.

1. INTRODUCTION

Music is split into various genres affected by culture and region. The genres also depend on ages. A previous study concerning the relationships between music and ages conducted a keyword analysis using lyrics. Since keywords were by and large subjectively selected, the study did not objectively discuss the evolution of music. Ideally, an analysis based on more objective parameters, such as MIR parameters, is desirable. Another previous study investigated relationships between music and ages in western popular music. One problem with the study is that music that covers several genres may have affected the results. Another problem is that it could not calculate additional MIR parameters, because musical audio materials used in the study are not provided. These problems are solved by collecting musical audio and selecting music genres in order to estimate the change in music across ages objectively.

Therefore, this study investigated the change in J-anime music (JAM) over several decades from musical audio. JAM is selected in order that a specific musical genre can be focused on. Specifically, 618 musical audio of JAM are collected, and a set of 25 types of MIR parameters are calculated from them. Then the values of the parameters for tunes in different decades are evaluated. Then the MIR parameters are used to classify the tunes by the *k*-means algorithm, which is known as an unsupervised clustering method. The unsupervised clustering is done so that we may obtain objective clusters of the 618 audio using the 25 types of parameters. It may show the performance of the 25 parameters by describing the characteristics of the audio's age. Then we conducted a decade estimation using the Naïve Bayes and Random Forest methods under a 10-fold cross validation scheme. Released decades were then estimated using each of the 25 parameters so as to confirm the effectiveness of each parameter as well. By using the results of the decade estimation using a single parameter, the study squeezes the set of parameters into a set of fewer parameters. In addition, the study discusses the correspondence relationships between released decades of JAM and MIR parameters of musical audio using mean values of each parameter across decades. Finally, released decades are subjectively evaluated by ten subjects who are asked to guess the decade released for 60 songs of JAM. Then, the study confirms the performance of automatic estimation of released decades by comparing the results of objective automatic estimation and subjective evaluation. This study may contribute to realizing an automatic estimator for released decades from musical audio.

2. J-ANIME MUSIC

2.1. Definition of J-Anime Music

Animation, put simply, involves making a motion picture by using static images. This is done by cameras or video processing software for film shooting using drawings, clay figures, and others. On the other hand, the Japanese term *anime*, which is derived from "animation", mainly means hand-drawn or digital animation. These tell stories and are produced for commercial proposes as movies, television programs, and so forth in Japan. For Japanese people, animation is *anime*. In this study, J-anime music (JAM) is defined as follows: songs used as theme songs or soundtracks in J-anime, songs sung by voice actors who voice roles in J-anime but not used in J-anime, songs composed on the basis of J-anime stories, or songs representing the image of current J-anime.

2.2. Evolution of J-Anime

Changes in J-anime are said to have been caused by the domestic impact of Japan's rapid economic growth. For example, many J-anime from the 1960s to the 1980s were science fiction or had heroes or heroines generated by hand-drawn animation. On the other hand, since the 1990s, exported J-anime to other countries was created by reducing production cost and efficient animation production due to the change of production techniques. Therefore, the contents of J-anime have been changing. Such changes in J-anime may cause changes in JAM, but no study has so far investigated this. Examples of popular J-anime from the 1960s to now are shown in Table 1.

Table 1: Example of popular J-anime in each decade.

Decade	J-Anime title	J-Anime Genre
1960	Astro Boy (鉄腕アトム), GeGeGe no Kitarō (ゲゲゲの鬼太郎), Sasuke (サスケ)	Science fiction, Horror, Hero
1970	Chōdenji Robo Combattler V (超電磁ロボ コン・バトラーV), Cutie Honey (キューティーハニー), A Dog of Flanders (フランダースの犬)	Mecha, Magical girl, masterpiece series
1980	Urusei Yatsura (うる星やつら), Fist of the North Star (北斗の拳), Kinnikuman (キン肉マン)	Comedy, Action, Battle
1990	Magical Circle Guru Guru (魔法陣グルグル), Cardcaptor Sakura (カードキャプターさくら), Blue Submarine No. 6 (青の6号)	Adventure, Magical girl, Science fiction
2000	Gravion Zwei (超重神グラヴィオンZwei), Pokémon: Advanced Generation (ポケットモンスター アドバンスジェネレーション), Code Geass: Lelouch of the Rebellion (コードギアス 反逆のルルーシュ)	Mecha, Adventure, Alternate history
2010	K-On!! (けいおん!!), Angel Beats!, Attack on Titan (進撃の巨人)	Slice of life, Comedy, Dark fantasy

3. OUTLINE OF PROPOSED MEHTOD

3.1. Investigation of the Relationship between the Change in JAM and Values of 25 Parameters

To calculate the MIR parameters that represent the change in JAM over decades, this paper investigates the values of all 25 MIR parameters for tunes in different decades. Specifically, the 25 MIR parameters are firstly calculated by using the collected 618 musical audio of JAM. Next, we conducted an automatic estimation of decades by using the MIR parameters calculated for JAM. In addition, each parameter is evaluated in terms of the estimation ability of decades and some parameters confirmed as effective in the estimation are extracted to estimate the decade. In addition to collecting the 618 musical audio of JAM from the 1960s to the 2010s, we categorized them into each decade.

3.2. Calculation for MIR Parameters

Several MIR parameters are calculated for the following statistics: average (avg), standard deviation (sd), kurtosis (ku), and skewness (sk). The MIR parameters are as follows:

Spectral centroid is calculated by the ratio of total sum multiplied by power of current frequency power and the total sum to frequency power. Spectral centroid is known to represent a measure of impression of a sound's brightness.

Spectral flux is calculated by the amount of power change in frequency spectrum between conjunct two frames.

Spectral flatness is a measure of the flatness of power spectrum, calculated by dividing geometric mean by arithmetic mean of power of all subbands.

Spectral rolloff is a measure of the frequency that contains a certain percentage of the power spectrum (.95 in this study) from low frequency.

Spectral Bandwidth is the difference between the upper and lower frequencies in a continuous set of frequencies. If several musical instruments are played in the spectral bandwidth, it is likely to have a large value, and vice versa.

Root mean square (RMS) is a measure of acoustic intensity.

Average tempo is a measure of the average interval between beats.

3.3. Relationship between MIR Parameters and Decades

Figures 1, 2, 3, and 4 show the averages of each decade in terms of sk of spectral rolloff, avg of spectral flatness, RMS, and bandwidth, respectively. Error bars in figures 1-4 show standard deviation, and horizontal axes indicate released decades. Figure 1 shows the average sk of spectral rolloff has decreased every decade. Figures 2 and 3 show the average avg of spectral flatness and RMS has increased every decade. Figure 4 shows the avg bandwidth has a salient difference between 1960s and the other decades.

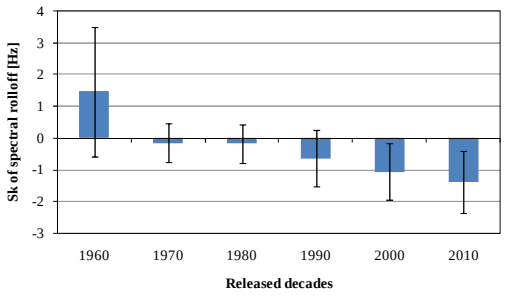


Figure 1: Average sk of spectral rolloff of each decade.

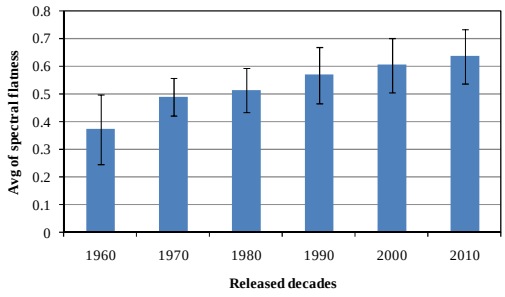


Figure 2: Average avg of spectral flatness of each decade.

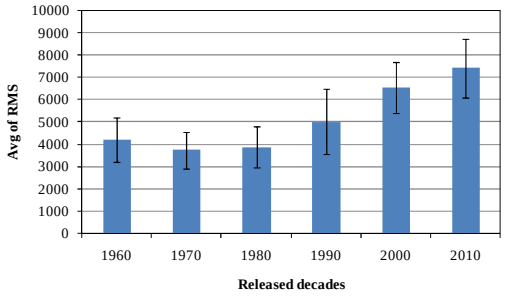


Figure 3: Average avg of RMS of each decade.

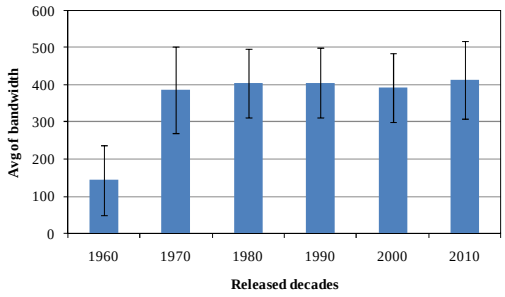


Figure 4: Average avg of bandwidth of each decade.

4. ESTIMATION OF RELEASED DECADES

4.1. Outline of Clustering Method

The validity of the 25 parameters to estimate the decade of the tunes in JAM is investigated by conducting a widely-used *k*-means clustering of unsupervised clustering. In this way, which cluster belongs to which is investigated. In addition, to estimate released decades of JAM, a Naïve Bayes classifier, which is a supervised identification machine learning based on Bayes' theorem, and a Random Forest, which is an algorithm of supervised machine learning for classification and estimation, are employed to estimate the decade of the tunes in JAM. These two types of machine learning are employed under the 10-fold cross validation scheme. In this study, a decision tree is built constituted by the calculated MIR parameters as a condition on each branch.

Results for the clustering concern the rate of correct recognition and evaluation index of reliability, obtained by the kappa statistic that shows the extent of the coincidence with a confusion matrix that shows all of the samples as classification results.

4.2. Result of *k*-means Clustering

Figures 5 and 6 show clustering results when K=6 and K=3, respectively. In addition, figure 7 shows clustering results for K=3 for two-decade periods. By using the 618 musical audio of JAM calculated by 25 MIR parameters, figure 5 shows the following clustering results: cluster 6 is mainly for 1960s, cluster 3 is mainly for 1970s, cluster 5 is for 1980s, several clusters are for the 1990s, and clusters 1, 2, and 4 are for the 2000s and the 2010s. Furthermore, the clustering results in figure 6 show the following: cluster 3 is for the 1960s, cluster 2 is for both the 1970s and the 1980s, and cluster 1 is for from the 1990s to the 2010s. Moreover, the clustering results in figure 7 show the following: cluster 3 is for the 1960s and 1970s, cluster 2 is for the 1980s and the 1990s, and cluster 1 is for the 2000s and the 2010s. Therefore, plural clusters are mostly divided between decades. Accordingly, when cluster numbers are K=3 and K=6, it is possible to classify the tunes into decades, which

implies that the 25 parameters can describe the released decades of the tunes of JAM. For this reason, we estimate the decades by classifying tunes into single or double decades from the 1960s to the 2010s.

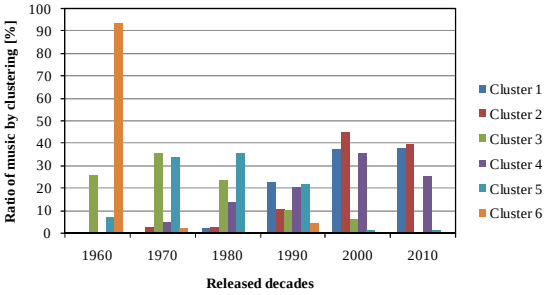


Figure 5: Clustering results for decades when *K*=6.

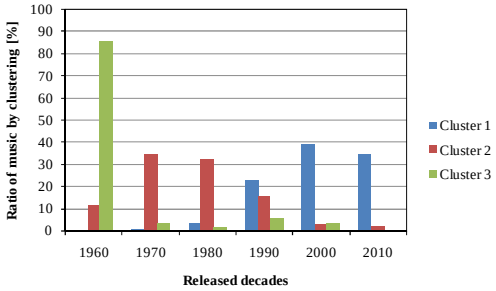


Figure 6: Clustering results for decades when *K*=3.

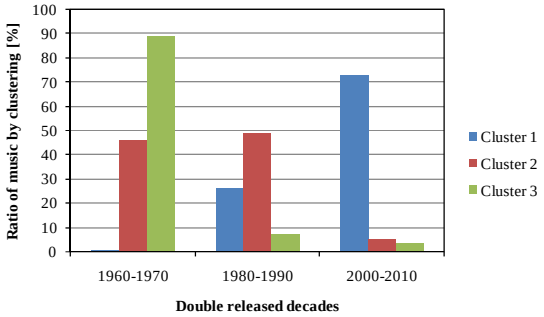


Figure 7: Clustering results for double decades when *K*=3.

4.3. Results by Naïve Bayes and Random Forest

The present study estimates the released decades of JAM by Naïve Bayes and Random Forest using the 25 MIR parameters.

4.3.1. Results by Naïve Bayes

In estimation results of released decades by Naïve Bayes, the overall accuracy rate is 52.75% and kappa statistic are .43 for single decades. In addition, accuracy rates are 70.39% and kappa statistics are .55 for estimating double decades. The results of single and double decades estimated by Naïve Bayes are shown in tables 2 and 3, respectively. Tables 2 and 3 show estimated decades vs. correct decades, accuracy rates of correct estimation, the number of songs estimated correctly, and the number of misidentified tunes. In table 2, the 1960s and 1980s have the highest and lowest accuracy rates. Furthermore, in table 3, the 1960s to the 1970s and the 2000s to the 2010s have higher accuracy rates than 1980s-1990s.

		Estimated decades					
		1960	1970	1980	1990	2000	2010
Correct decades	1960	61	15	6	1	0	0
	1970	3	61	39	6	1	0
	1980	2	39	56	6	7	0
	1990	4	15	19	33	29	10
	2000	2	2	6	9	51	40
	2010	0	1	5	1	24	64
Accuracy rate		84.72%	45.86%	42.75%	58.93%	45.54%	56.14%

Table 2: Results of single decades estimated by Naïve Bayes.

		Estimated decades		
		1960-1970	1980-1990	2000-2010
Correct decades	1960-1970	105	87	1
	1980-1990	24	152	44
	2000-2010	5	22	178
Accuracy rate		78.36%	58.24%	79.82%

Table 3: Results of double decades estimated by Naïve Bayes.

4.3.2. Results by Random Forest

In estimation results of released decades by Random Forest, the overall accuracy rate is 54.37% and kappa statistic is .45 for single decades. In addition, the accuracy rate is 73.30% and kappa statistic is .60 for double decades. Results of estimated single and double decades by Random

Forest are shown in tables 4 and 5, respectively. Tables 4 and 5 show estimated decades vs. correct decades, accuracy rates of correct estimate, the number of songs estimated correctly, and the number of misidentified tunes. In table 4, the 1960s and 1980s have the highest and lowest accuracy rates. Furthermore, in table 5, the 2000s to the 2010s have the highest accuracy rate.

		Estimated decades					
		1960	1970	1980	1990	2000	2010
Correct decades	1960	71	8	3	1	0	0
	1970	6	65	30	7	2	0
	1980	3	39	47	15	6	0
	1990	1	12	30	34	26	7
	2000	1	4	10	13	62	20
	2010	0	0	4	10	24	57
Accuracy rate		86.59%	50.78%	37.90%	42.50%	51.67%	67.86%

Table 4: Results of single decades estimated by Random Forest.

		Estimated decades		
		1960-1970	1980-1990	2000-2010
Correct decades	1960-1970	149	42	2
	1980-1990	43	142	35
	2000-2010	4	39	162
Accuracy rate		76.02%	63.68%	81.41%

Table 5: Results of double decades estimated by Random Forest.

4.3.3. The Validity of MIR Parameters

From the results of the estimation for single decades, the 1960s shows high accuracy rate whereas the 1980s shows low accuracy rates when using Naïve Bayes or Random Forest. Moreover, when estimating double decades, 2000s to 2010s shows a high accuracy rates when using Naïve Bayes or Random Forest. Therefore, in the case of single decades, MIR parameters to characterize JAM in the 1960s could be found. In addition, in the case of double decades, MIR parameters to characterize JAM from the 2000s to the 2010s could be found. However, to distinguish JAM in the 1980s by Naïve Bayes is difficult since songs from the 1980s are likely to be classified into the range of 1970s or 1990s. Furthermore, JAM in the 2000s and 2010s may have similar MIR parameters. On the other hand, from the results of Random Forest, to describe JAM in the 1980s is difficult because songs from the 1980s are likely to be identified into any decade from 1970s to 2000s, as shown in table 4.

4.4. Estimation Result of Release Decades by MIR Parameters

In short, the decades after 1970s are difficult to distinguish correctly. To investigate the validity of MIR parameters in each decade, we try to estimate the decades by using an MIR parameter. Accordingly, we extract a certain amount of parameters that definitely contribute to the decade estimation. Meanwhile, released decades of JAM are then estimated by a squeezed MIR parameter set. Results of estimated decades using each single MIR parameter are shown in table 6.

Next we will explain our method to squeeze the parameter set. Firstly, we obtained the accuracy rates when estimating single or double decades. Then the parameter that provides the lowest score is chosen and tentatively got rid of from the parameter list. Then the single or double decades is estimated using the remaining parameters. If the accuracy rate grows, we try to find the lowest from the remaining parameters and then do the same steps until the improvement of the accuracy rate disappears. As a result, the selected MIR parameter set in decades by Naïve Bayes is 21 parameters, excluding ku of spectral flux and sd, ku, sk of bandwidth. In single decades by Random Forest, the set is 24 parameters, excluding sk of bandwidth. Moreover, in double decades by Naïve Bayes and Random Forest, the set is 25 parameters without change.

		Accuracy rate [%]			
		Naïve Bayes		Random Forest	
		Single decades	Double decades	Single decades	Double decades
Spectral Centroid	avg	35.92	55.18	29.29	44.34
	sd	26.05	33.66	18.93	33.50
	ku	25.08	42.72	19.90	36.89
	sk	30.58	51.62	23.46	36.89
Spectral Flux	avg	37.86	58.58	28.16	51.29
	sd	29.61	48.54	18.12	41.42
	ku	17.31	39.16	18.93	30.91
	sk	32.36	57.61	24.43	46.28
Spectral Flatness	avg	38.03	54.69	25.08	42.07
	sd	30.91	46.76	21.04	35.76
	ku	29.13	44.82	21.52	35.76
	sk	42.23	58.09	36.08	57.61
Spectral Rolloff	avg	37.54	55.02	27.83	46.44
	sd	23.62	37.70	17.64	33.17
	ku	26.86	44.66	21.36	35.60
	sk	33.98	53.72	27.18	42.23
Bandwidth	avg	29.61	49.03	29.29	43.20
	sd	19.74	38.51	21.20	35.92
	ku	19.26	39.16	17.64	37.70
	sk	21.52	39.32	17.15	36.41
RMS	avg	36.41	59.06	26.38	48.06
	sd	37.54	57.12	30.74	51.13
	ku	33.17	56.63	28.32	53.24
	sk	35.44	59.22	23.79	49.51
Average Tempo		24.43	39.97	24.60	37.06

Table 6: Results of estimation of decades using each single MIR parameters.

4.5. Re-estimated Results of Released Decades By Squeezed Parameters

This subsection describes the estimation of released decades of JAM by Naïve Bayes and Random Forest using the selected MIR parameter set.

4.5.1. Re-estimation Results of Investigation by Naïve Bayes

In estimation results of released decades by Naïve Bayes, the accuracy rate is 55.66% and kappa statistic is .46 for single decades. Moreover, compared with the results by the 25 parameters, the accuracy rate for decades is increased in 2.91 points. Results of re-estimated single decades by Naïve Bayes are shown in table 7, where the 1960s has the highest accuracy rate. Furthermore, the 1970s, the 1980s, and 1990s still have low accuracy rates. In addition, compared with the accuracy rate in table 3, accuracy rates in the 1970s and after are increased. Therefore, we confirmed that the selected MIR parameter set is better than the unselected 25 MIR parameters.

		Estimated decades					
		1960	1970	1980	1990	2000	2010
Correct decades	1960	60	17	4	2	0	0
	1970	4	68	31	6	1	0
	1980	2	35	59	7	7	0
	1990	3	15	19	35	29	9
	2000	2	2	5	7	57	37
	2010	0	2	3	2	23	65
Accuracy rate		84.51%	48.92%	48.76%	59.32%	48.72%	58.56%

Table 7: Results of single decades re-estimated by Naïve Bayes.

4.5.2. Re-estimation Results of Investigation by Random Forest

In estimation results of released decades by Random Forest, the accuracy rate is 55.50% and kappa statistic is .46 for single decades. Moreover, compared with those by the 25 parameters, the accuracy rate for single decades is increased in 1.13 points. Results of re-estimated decades for decades by Random Forest are shown in table 8, where the 1960s has the highest accuracy rate. Furthermore, compared with the accuracy rate in table 4, accuracy rates of those other than the 1970s and the 2000s are increased. Therefore, the selected MIR parameter set is better than the unselected 25 MIR parameters.

		Estimated decades					
		1960	1970	1980	1990	2000	2010
Correct decades	1960	72	7	4	0	0	0
	1970	4	63	36	4	3	0
	1980	3	37	51	13	5	1
	1990	1	15	20	41	28	5
	2000	1	3	8	23	60	15
	2010	0	3	1	5	30	56
Accuracy rate		88.89%	49.22%	42.50%	47.67%	47.62%	72.73%

Table 8: Result of single decades re-estimated by Random Forest.

5. SUBJECTIVE EVALUATION OF DECADES

5.1. Subjective Evaluation Experiment of Released Decades of JAM

As shown above, the selected MIR parameters are definitely effective for estimating released decades of JAM by both Naïve Bayes and Random Forest. However, no relationship is confirmed between estimated results of machine learning and people's impressions. Therefore, an investigation is required into impressions of released decades given in the music. In this section, the authors describe a subjective evaluation experiment to confirm the effectiveness of the selected MIR parameter set.

5.2. Outline

A subjective evaluation experiment was conducted, where subjects were asked to guess decades while listening to JAM tunes using personal computers with headphones in a quiet room. The sound pressure level was freely chosen by each listener. This experiment employed 60 musical audio of JAM, and participants were ten university students (eight males and two females), whose average age was 21.60 (sd=1.08).

5.3. Method

First, the subjects were asked to listen to the first 15 seconds of each musical audio of JAM. In addition, subjects voluntarily listened to a free section of musical audio. Second, subjects were allowed to listen to musical audio again and then selected any one of the released decades from the 1960s to the 2010s. In addition, they were asked to answer whether or not they had previously listened to the tune.

5.4. Results

Accuracy rates of each subject are shown in figure 8, where the horizontal axis shows the index of each subject and the vertical axis shows the accuracy rate. In the results of the subjective evaluation experiment, accuracy rates vary across subjects. In short, the ability of people to guess released decades is above 33.33% accuracy for single decades and 66.67% accuracy for double decades.

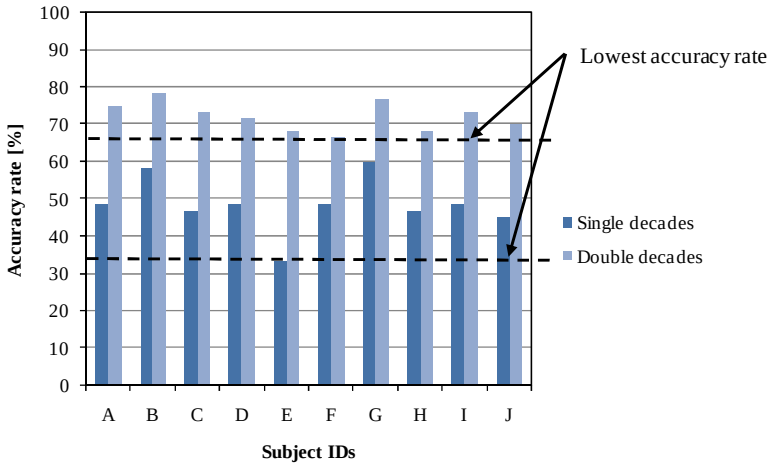


Figure 8: Accuracy rate of each subject.

5.5. Discussion

The performance of the selected MIR parameter set is confirmed. Results of estimation and experiment are shown in figures 9 and 10, where estimated results in single decades by Naïve Bayes and Random Forest are 3rd and 4th among the ten subjects, respectively. On the other hand, estimated results in double decades by Naïve Bayes and Random Forest are 8th and 6th among the ten subjects, respectively. Therefore, results of the subjective evaluation experiment and objective estimation do not show any salient difference in either single or double decades. Moreover, the two methods performed comparably to people.

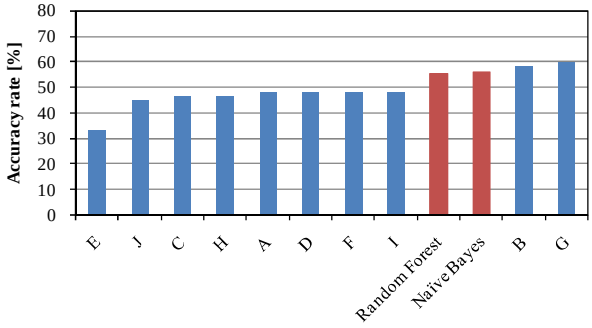


Figure 9: Results of estimation and experiment for single decades.

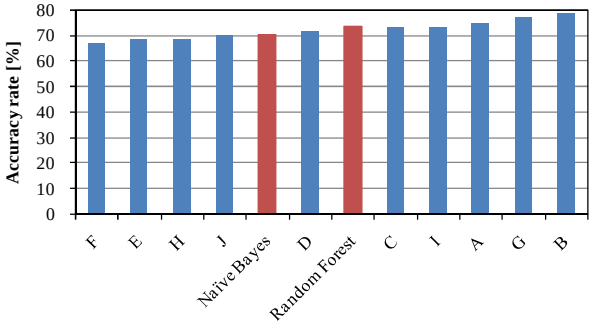


Figure 10: Results of estimation and experiment for double decades.

6. CONCLUSION

The change in Japanese anime music (JAM) over decades was investigated by using musical audio to extract the MIR parameters. Specifically, the musical audio of the JAM were firstly collected, and then 618 musical audio were employed. In addition, 25 types of MIR parameters were calculated from the collected musical audio. The MIR parameters were used to classify tunes on the basis of the *k*-means algorithm for unsupervised clustering. Released decades of each tune were then estimated by using the parameters, and the effectiveness of each parameter for decade estimation was discussed. Decades were estimated by Naïve Bayes and Random Forest under a 10-fold cross validation scheme. Then the present study estimated the released decades using a set of squeezed parameters. Moreover, this study discussed the

correspondence relationships of released decades of JAM and MIR parameters of musical audio on the basis of the re-estimated result. Finally, a subjective evaluation of released decades was done, in which ten subjects were asked to guess the decade released for 60 songs of JAM. Furthermore, this study confirmed the effectiveness of automatic estimation of released decades from the results of estimation using machine learning and the subjective evaluation experiment. Therefore, the authors confirmed that the proposed method has the potential to estimate the decades more accurately than people. For future work, we investigate MIR parameters that are able to describe the decades being difficult to estimate.

7. ACKNOWLEDGMENTS

This study is partly supported by the Grants-in-Aid for Scientific Research (25580050).

8. REFERENCES

Otsuka, M. (2004). Changes in Some Keywords and the Concept of Love in Modern Japanese Pop Songs. Language and Culture, 17, *30-66. (in Japanese)*

Serrà, J., & Corral, À., & Boguñà, M.,& Haro, M., & Arcos, J. Ll. (2012). Measuring the evolution of contemporary western popular music. Scientific reports, 2, *1-6.*

Takahashi, M., & Tsugata, N. (2011). Anime Studies. NTT Publishing, *3-5. (in Japanese)*

Yamaguchi, Y. (2009). Nihon no anime zenshi: sekai o seishita Nihon anime no kiseki. TEN-BOOKS, *1-223. (in Japanese)*

Tzanetakis, G., & Cook, P. (2002). Musical Genre Classification of Audio Signals. IEEE Transactions on speech and audio processing, 10, 5, *293-302.*

Yang, Y. H., & Chen, H. H. (2011). Music Emotion Recognition. CRC Press, *42-44.*

Hiraga, Y., & Ohishi, Y., & Takeda, K. (2009). Method for Calculating the Subjective-based Music Similarity Measure. Information Processing Society of Japan, 2, *1-6. (in Japanese)*

Hirae, R., & Nishi, T. (2008). Mood classification of music audio signals. Acoustical Society of Japan, 64(10), *607-615. (in Japanese)*

Ono, Y., & Katto, J. (2009). An Automatic Estimation Method of Music Emotion Based on Musical Psychology. PSJ SIG Technical Report, 23, *61-66. (in Japanese)*

[3-04] AN EXPERIMENTAL STUDY ABOUT BEHAVIORAL SYNCHRONY BETWEEN CHIMPANZEES

Lira Yu[1], Masaki Tomonaga[1]
[1]Primate Research Institute, Kyoto University, Japan

Poster

Humans often unconsciously synchronize the movements and its timing of others. Examples includes hand-clapping in a concert hall and walking along the street with others. This timing match is thought to have socially adaptive functions and to be shared with other highly social non-human primate species. There are some reports that chimpanzees, who are the member of our closest living relatives, show vocal chorusing. However few studies have investigated the nature of behavioral synchrony in chimpanzees under well controlled laboratory settings. Here we examined a pair of chimpanzees in two different experimental settings: side-by-side and face-to-face settings. A finger-tapping task was introduced to produce their preferred rhythmic movements. In side-by-side setting, auditory information of the movements was exchanged between chimpanzees. On the other hand, in face-to-face setting, auditory and visual information of the movements were exchanged between chimpanzees. We found that chimpanzees matched the movements in time with others in face-to-face setting more significantly than in side-by-side setting. This result demonstrates that visual information of the other's movement is required to facilitate synchronization between chimpanzees.

[3-05] MUSIC MODULATES THE STRENGTH OF ILLUSORY SELF-MOTION PERCEPTION (VECTION).

Takeharu Seno[1]
[1]Kyushu University, Japan

Poster

Exposure to a visual motion field that simulates the retinal optical flow generated by self-movement commonly causes the perception of the subjective movement of one's own body. This phenomenon is known as 'vection'. We presented four types of music (two fast and two slow tempo music) during illusory self-motion perception (vection). Vection was induced in fourteen stationary observers by presenting expanding optic flow. Optic flow displays (72° × 57°; presented for 30s) consisted of 16,000 randomly positioned dots with global dot motion to simulate forward self-motion (16m/s). Participants were asked to press a button when they perceived forward self-motion, and keep the button depressed for the duration of self-motion. After each trial, the participants rated subjective vection strength using a 101-point rating scale ranging from 0 (no vection) to 100 (very strong vection). There were five music conditions: two fast, two slow, and a no-music condition. These conditions were conducted in separate sessions. Three trials were conducted for the with-music conditions and four trials for the without-music condition. For the two fast music conditions, we used "Bakushou-sengen" (the theme of a Japanese Pro-wrestler) and the theme music from the movie "Back to the Future", and for two slow music conditions, we used Pachelbel's "Kanon-D-dur" and Debussy's "Moon Light". The three trials were conducted successively within the same session.
The average latencies and durations were shortest and longest respectively in the two fast music conditions, shorter in the two slow music conditions, and slowest in the no-music condition. The average magnitudes were largest in the two fast music conditions and the Kanon-D-dur condition, and smallest in the Moon Light and no-music conditions. Vection was facilitated by two fast and one slow tempo music rather than without music condition. We speculated that the fast tempo and the active music might induce higher arousal level in the participants rather than without music condition and the higher arousal level might induce stronger vection.

[3-06] GESTURAL CONTENT INFLUENCES EVALUATIONS OF ENSEMBLE PERFORMANCE

Steven Morrison[1], Harry Price[2], Eric Smedley[3], Cory Meals[1]
[1]School of Music, University of Washington, USA, [2]School of Music, Kennesaw State University, USA, [3]Jacobs School of Music, Indiana University, USA

Poster

Audience evaluations of ensemble performances vary depending on the expressivity of conductor gestures, even among identical performances. Here we test whether varying specific constituent information will affect listeners' assessment of specific performance parameters. Isolating articulation and dynamics, we test whether the degree of gestural variability within these two areas affects the way in which they are perceived. In other words, would effects observed at the broad level of "expressivity" persist among two of its specific constituent parts? Four music excerpts—two featuring contrast in articulation content and two featuring contrast in dynamic content—were performed by a large chamber ensemble. Each excerpt was prepared and recorded in both low- or high-contrast conditions specific to their performance parameter (articulation or dynamics). Performances were then paired with video of one of four conductors conducting with either a high or low degree of gestural expressivity appropriate to the target parameter. Using two equivalent test forms, college music majors and nonmajors (N = 288) viewed sixteen 30-second performances and evaluated ensemble articulation, dynamics, technique, tempo and overall expressivity.Evaluations of performance expressivity were highly correlated with evaluations of articulation and dynamics (r = .71 and .85, respectively). Articulation and dynamics scores alone predicted a full 74.5% of expressivity variance. This suggests that broad judgments of performance quality may be affected by variation in even one musical parameter. Likewise, strong correlations between evaluations for the target characteristic and the alternate (non-varied) characteristic (r = .77 for dynamics items, r = .68 for articulation items) indicated that participants found it difficult to discriminate between multiple specific aspects of ensemble performances. The addition of gestural variability along one musical dimension was associated with more positive evaluations of performance along that dimension, as well as evaluations along an unrelated dimension and overall expressivity, even where the ensemble performance was invariant. Results lend support to the hypothesis that visual information plays a critical role in evaluation of music performance. Further, these data suggest that even highly specific musical/visual information can have a "spill-over" effect across perception of varied musical parameters as well as overall assessment of musical expression.

[3-07] THAT'S THE WAY I LIKE IT: RELATIONSHIPS BETWEEN MUSICAL GENRE AND DANCE PREFERENCES AND MUSIC-INDUCED MOVEMENT

Birgitta Burger[1], Suvi Saarikallio[1], Marc Thompson[1], Geoff Luck[1], Petri Toiviainen[1]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland

Poster

Listening to music makes us move. Music-intrinsic features, for instance beat strength, pulse clarity, or emotional content of music, have been shown to influence music-induced movements. Moreover, individual factors, such as personality, affect movability to music. Additionally, music-induced movement may be shaped by preference for dancing and for certain musical genres. This study aims at investigating how musical genre and dance preferences relate to music-induced movement. It was assumed that participants with a preference for dance music genres, such as Techno, exhibit higher amounts of movement than participants with a preference for less danceable genres, such as Jazz. Furthermore, it was hypothesized that participants who like to dance exhibit higher amounts of movement. Sixty participants were presented with 30 musical stimuli representing different genres of popular music. Participants were asked to move along in a natural way. An optical motion capture system was used to record participants' movements, from which 17 movement features related to speed, acceleration, complexity, amount, and rotation were subsequently extracted. Subsequent to the movement part, participants answered to a questionnaire related to their music and dance preferences. Results indicate that participants with preferences for Pop and Techno exhibited higher amounts of movement than participants disliking these genres. Furthermore, a U-shaped relationship was found for preference for Latin music: participants either liking or disliking Latin moved more than other participants. Participants with dance as hobby moved considerably more than participants without dance as hobby. Additionally, participants who reported that they like and often go out to dance, showed higher movement activity during the experiment than participants who less liked dancing. This study revealed relationships between music-induced movements and preferences for musical genres and dancing. Participants liking Pop and Techno might be more inclined to dance to music, as such genres are more related to dancing (e.g., being played in nightclubs) than, for example, Jazz. The result that more movement could be detected for participants, who liked to dance, supports our hypothesis and suggests a relationship between preference for dance and movement exhibited. This finding is to be considered when designing dance-related experiments in future research.

[3-08] TEMPO ANALYSYS OF DANCE MOVEMENTS: THE MASTER AND A DISCIPLE OF NIHON BUYO

Hiroshi Kawakami[1], Yuki Mito[1], Toru Ozawa[1], Yukitaka Shinoda[2], Toshihiro Irie[2], Mieko Marumo[1]
[1]College of Art, Nihon University, Japan, [2]College of Science and Technology, Nihon University, Japan

Poster

1. Background: Nihon Buyo, that is one of Japanese traditional dance, needs high-level technique. These dance music is called Nagauta which have the peculiar tempo. Therefore, the marked difference in its expression can be felt between a master and disciples. Some knowledge has been acquired to the motion of okuri that is fundamental operation of Nihon Buyo. Moreover, some analysis methods have also been considered, such as the method using Laban Movement Analysis or Motion Averaging. However, these researches have not completely come to find out the feature of experts yet.

2. Aim: In this research, the dance of a living national treasure was analyzed. A dance is a successive change of stillness and movement and this change determined the contrast of its impression. Then the motion tempo of dancer was investigated to find out this change of stillness.

3. Method: Optical motion capture system was used to acquire the movements of dancers. This system has 12 cameras, capturing 42 markers on a body at a frame rate of 60 fps. The master and his disciple danced a part of Hichifukujin. The relative movement from a sacrum was calculated on 5 places, which were on a head, both hands and both toes, using these motion data. And these movements were converted to tempo data.

4. Result: As the maximum tempo of each choreography was compared by the paired t-test, maximum values of the master had exceeded the disciple significantly on both hands and toes ($p<0.05$). Moreover, in a viewpoint of the envelope on the time variation, we found the difference on the way of tempo change. Especially, at the quick choreography, the initial tempo of the master had an inclination to increase rapidly.

5. Conclusions: In this research, although change of dance tempo was considered, remarkable difference could be found out between the master and his disciple. Especially, as we found out the remarkable change of motion on the master's data, this element may have affected visual

impressions. Since the liveliness of dance motion related to this tempo change surely, the validity of investigating tempo change was suggested by this research.

[3-09] BETWEEN PEAKS AND TROUGHS: LOCAL AND GLOBAL ASPECTS OF CONTOUR AND THE PERCEPTION OF MELODIC DIRECTION

Neta Cohen-Shani[1,2], Zohar Eitan[2], Dominique Lami[1,2]
[1]Tel Aviv University, School of Music, Israel, [2]Tel Aviv University, Department of Psychology, Israel
Poster

Navon's hierarchical letters (1977) -- larger letter shapes composed of congruent or incongruent smaller letters -- have been widely employed to investigate whether global or local aspects of the visual input enjoy processing dominance. Recent research has examined auditory parallels to Navon's paradigm. Initially, Justus & List (2005) presented a 9-tone sequence composed of three-note chunks and manipulated global and local dimensions (i.e., pitch contour within chunks vs. contour between them) independently. Studies employing such stimuli, using both neurophysiological (ERP) and speeded behavioral methods, have demonstrated global precedence, measured as earlier responses to global features, in melodic processing. The above studies, however, insufficiently address potentially crucial stimuli and population variables --- particularly, global and local pitch intervals as well as musical expertise. Furthermore, the premise that response precedence necessarily entails processing dominance is not evident when stimuli are themselves time-ordered sequences. Hence, in this study global vs. local predominance was assessed based on participants' subjective (and unspeeded) evaluation of pitch direction, rather than on temporal precedence in speeded response. Furthermore, we systematically manipulated pitch interval and pitch direction globally and locally, and examined both musicians and non-musicians. 24 participants (12 musicians) were presented with sixteen 900 ms. 3X3 tone sequences. Pitch interval and direction were manipulated within and between chunks. Subjects rated on a discrete scale [(-4)-(+4)] to what degree they subjectively perceived each sequence as falling or rising in pitch. Consistent with previous research, ratings were mostly affected by global features. In conditions highlighting local grouping (e.g., narrow within-chunk intervals) local effects were also observed. There was little effect of expertise. Importantly, a highly significant correlation (Spearman r=0.94) was found between participants' ratings and the time interval separating the lowest and highest tones. This suggests that contour processing may be largely based on detecting and comparing melodic peaks and troughs, regardless of the finer contour patterns. Thus, while contour processing may indeed display global predominance, such predominance may itself reflect processing of conspicuous local details. This finding calls for reassessment of the notion of global predominance in auditory contexts, problematizing attempts to create auditory parallels to Navon's paradigm.

[3-10] DIFFERENCES IN BRAIN RESPONSES TO MELODIC AND RHYTHMIC DEVIATIONS IN MUSICIANS

Claudia Lappe[1], Markus Lappe[2], Christo Pantev[1]
[1]Institute for Biomagnetism and Biosignalanalysis, University Clinic Münster, Germany, [2]Institute for Psychology, University of Münster, Germany
Poster

1. Background: Perception and learning theories state that the brain generates predictions about future events based on statistical sensory learning. This phenomenon offers a promising approach to understand music perception. Humans are able to predict upcoming musical events based on previously learned musical regularities. To investigate predictive mechanisms in the human brain researches have used electrophysiological markers for prediction violations like the mismatch negativity (MMN), an event related component that occurs automatically when an auditory stimulus is different in pitch, loudness, timbre or tempo. Two previous magnetoencephalography (MEG) studies with musical inept subjects demonstrated that the MMNs to melodic and rhythmic deviations increase after short-term musical training showing that acquired musical expertise leads to better predictions about upcoming musical events.

2. Aim: In the present study we asked whether prediction violation in the pitch and rhythm domain are processed similarly in professional musicians.

3. Methods: We recorded MEG responses in pianists listening to rhythmic and melodic deviations in musical material (a short 6-tone melody composed of a C-major- followed by a G-major broken chord). In the deviant pitch condition the last tone was lowered by a minor third. In the rhythm deviation condition the last tone appeared 100 ms earlier than in the standard melody. Furthermore, in order to see whether neural activation within a musical context is different from neural activation to a simple tone deviation we presented the same pitch and timing deviation within a simple sequence of identical tones.

4. Results: A clear MMN was found in all conditions, but its temporal development differed between conditions. The time course of the rhythmically elicited MMN within the musical sequence was similar to the MMN in both simple tone deviation condition. The MMN elicited by melodic deviations, however, appeared about 50 ms later.

5. Conclusions: The results suggest that the automatic detection of melodic deviations in musicians takes more time than the detection of simple pitch or rhythm deviations. This is consistent with the idea that predictions about melodic continuation are more complex than predictions about simple tone sequences, and indicate that they may involve different cortical networks.

[3-11] MUSICAL FEATURE PROCESSING IN THE BRAIN DURING A NATURALISTIC, CONTINUOUS LISTENING – A REPLICATION STUDY

Iballa Burunat[1,2], Petri Toiviainen[1], Vinoo Alluri[1], Brigitte Bogert[3], Tapani Ristaniemi[2], Mikko Sams[3], Asoke Nandi[2,5], Elvira Brattico[3,4]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland, [2]Department of Mathematical Information Technology, University of Jyväskylä, Finland, [3]Cognitive Brain Research Unit (cbru), Institute of Behavioral Sciences, University of Helsinki, Finland, Finland, [4]Brain & Mind Lab, Department of Biomedical Engineering and Computational Science (becs), Aalto University School of Science, Finland, Finland, [5]Department of Electrical & Computational Engineering, Brunel University, Uxbridge, United Kingdom
Poster

1. Background: Musical processing of low-level (timbral) and high-level (tonal and rhythmical) musical features in the brain during a naturalistic listening using functional magnetic resonance imaging (fMRI) has been shown to elicit large-scale neural responses in cognitive, motor and limbic brain networks (Alluri et al., 2012).

2. Aims: We aimed to replicate previous findings regarding the functional neuroanatomy of music processing in the brain under a naturalistic musical listening by using identical methodological approach and a similar group of participants as used in Alluri et al. (2012).

3. Methods: Participants' brain responses were recorded using fMRI during a naturalistic continuous listening of a modern tango. These were voxelwise correlated against time series of a set of musical features computationally extracted from the music. The similarity of results obtained from the two datasets was assessed by correlating the respective spatial maps for each musical feature. Significance was estimated by means of permutation tests.

4. Results: All features displayed positive spatial correlations between the maps. However, only the low-level (timbral) features reached significance.

5. Conclusions: Overall the results could be replicated with varying degrees of significance: the functional topography underlying timbral processing was significantly similar in both datasets, whereas the functional topography representing tonal and rhythmical processing, although positively correlated, did not reach significance. The success in recruiting similar brain circuits in the processing of timbral features in two different population samples may evidence more universal, unconscious processing mechanisms for low-level acoustic features in the music. However, tonal and rhythmical features may respond to more cognitive, top-down level mechanisms associated with a larger participant-dependent variability, which may inhibit the ability to replicate previous findings. In conclusion, the present study supports and validates approaches of more ecological validity in the field of cognitive neuroscience of music, while it exposes the need to reconsider certain methodological approaches in the study of more cognitive phenomena with high inter-subject variability.

[3-12] MUSICAL TRAINING REFLECTED IN BETA-BAND ACTIVITY TO METER

Dahye Bae[1], June Sic Kim[2,3], Kyung Myun Lee[4], Chun Kee Chung[1,2,5,6], Suk Won Yi[7,8]
[1]Department of Brain and Cognitive Sciences, Seoul National University College of Natural Sciences, Seoul, Korea, Republic of, [2]Meg Center, Department of Neurosurgery, Seoul National University Hospital, Seoul, Korea, Republic of, [3]Sensory Organ Research Institute, Seoul National University, Seoul, Korea, Republic of, [4]Smart Humanity Convergence Center, Department of Transdisciplinary Studies, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea, Republic of, [5]Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Korea, Republic of, [6]Department of Neurosurgery, Seoul National University Hospital, Seoul, Korea, Republic of, [7]College of Music, Seoul National University, Seoul, Korea, Republic of, [8]Western Music Research Institute, Seoul National University, Seoul, Korea, Republic of

Poster

1. Background: When we listen to a series of tones, we perceive it by grouping. This is the concept of meter. A lot of studies about meter have been tried using a theoretical approach or behavioral experiments. We have studied meter using a systematical stimuli with a long unit. Beta-band activity is considered to be an important frequency band in the research of rhythm, because it is associated with movement, which is easily connected to rhythm in music. It is also related to consciousness such as active concentration.

2. Aims: The purpose of our research is to observe detection of meter in the human brain. We also examined between-group differences by comparing musicians to non-musicians.

3. Method: In the stimuli, the inducing section consisted of a regular accent adjusted by intensity. But the main section that was analyzed was composed of tones with the same pitch, intensity, and duration. During the magnetoencephalography (MEG) recording, each subject listened to a repeated sound stimuli with a silent documentary.

4. Results: The subjects listened to same tones in different ways. The tones were presented with the exact same properties, but the brain response was different from what was heard. By analyzing extracted beta-band activity, we could find some features referred to as meter. At some points, beta power increased more in musicians than in non-musicians. Overall, musicians' beta power was much greater than non-musicians'.

5. Conclusions: Now we are able to discuss rhythm through a neurophysiological approach. The human brain can find meter almost automatically, but music experts do it much better. These findings show that beta-band activity is increased by the level of musical training. In addition, beta-band activity seems to prove that people perceive rhythm associated with movement.

[3-13] THE IMPACT OF EMOTIONS AND FINGER STRETCH ON THE DYNAMICS OF PIANO MUSIC

Xiaoluan Liu[1]
[1]Department of Speech, Hearing and Phonetic Sciences, University College London, United Kingdom

Poster

1. Background: The relations between the dynamics of piano music and emotions are often examined by using music synthesis (cf. Juslin & Sloboda, 2013). Nevertheless, the results are often mixed due to the high individual variability in perceptual judgement of the emotions conveyed. Moreover, how wide the fingers are stretched could also affect the dynamics of piano music (Sandor, 1981), which so far has not been systematically investigated.

2. Aims: Hence, this study aims to use piano playing experiment to examine the impact of emotions and finger stretch on the dynamics of piano music.

3. Method: Five emotionally neutral pieces of music were composed for this study, with each piece corresponding to a different degree of finger stretch, ranging respectively from small (1-2, 2-3, 3-4, 4-5) to large (1-5). Six pianists were instructed to play each piece with four different emotions: Anger, fear, happiness and sadness according the fingerings provided on the scores. To obtain MIDI data (i.e., velocity which reflects dynamics) without having to use electronic keyboards, a PianoBar scanner (which turns any standard pianos into MIDI enabled pianos) was attached to the keyboard of a Bösendorfer grand piano. Data were processed using Matlab.

4. Results: Repeated measures ANOVAs show significant impact of emotions and finger stretch ($p<0.05$) on dynamics. Post-hoc Tukey HSD tests show anger and happiness have significantly ($p<0.01$) higher dynamics than fear and sadness, while the differences between groups (anger vs. happiness; fear vs. sadness) are non-significant. Moreover, for each emotion, large finger stretch has significantly higher ($p<0.05$) dynamics than small finger stretch, especially when 1-5 (strong-finger combination) is compared with 3-4 (weak-finger combination).

5. Conclusions: Using real piano playing experiment, this study firstly confirms the impact of emotions on the dynamics of piano music as reported in music perception literature; secondly, as speculated by previous literature, finger stretch is demonstrated to be an influential factor of dynamics especially when strong-finger combination is compared with weak-finger combination. The results can not only enhance our understandings of the relations between music and emotions with regards to dynamics, but also shed further light on human motor movement.

[3-14] MUSICAL AND TECHNICAL DEMANDS IN PIANO WORKS BY TWO BRAZILIAN COMPOSERS

Carmen Fregoneze[1]
[1]*Carmen Célia Fregoneze, Brazil*
Poster

Throughout the twentieth century, Brazil's musical life witnessed an explosion of musical styles following musical tendencies in vogue in Europe and in the United States of America. This phenomenon greatly influenced the eclectic span of art music and the diversity of musical languages used in this century's musical production. A variety of musical currents began to emerge; within each trend, many composers preferred to choose a path in which they could manifest their own musical thoughts and adhere to a compositional style. As composers absorbed contrasting creative trends, the repertoire of Brazilian piano music became large and diverse. This paper will focus on a small but significant sample of Brazilian piano music by composers Almeida Prado and Edino Krieger which enormously contributed to Music Performance. Each of these composers produced a large number of works which demonstrate their musical inventiveness, creativity, and versatility. Their outputs during a fruitful span of time presented great pluralism, stylistic variety, musical and technical complexities of all kinds, as well as challenges for the performer. A glimpse into the literature on these composers confirms that they are composers of great significance in the Brazilian musical scene, as can be confirmed by the available bibliography. Moreover, the music by these composers is colorful, stylistically and aesthetically varied and contrasting. They were influenced by cosmopolitan music and branched into extremely varied paths, exploring multiple aesthetic tendencies. However, the way in which they contextualized and explored known elements generated an entirely new group of musical works, recognizable through the combination of heterogeneous techniques. The works by Almeida Prado and Edino Krieger are characterized by intense contrast between the following musical elements: melody, harmony, rhythm, form, textures, timbre, tone, and also the projection of emotional states. These composers wrote works, borrowing a large gamut of musical material from different repertoires, and following tendencies in vogue during the time of their own compositional pinnacles. Their legacy significantly contributes to the contemporary Brazilian music, enlarging and enriching it.

[3-15] PROFICIENCY ESTIMATION FOR PIANO PERFORMANCE OF BURGMUELLER'S "LA CANDEUR"

Satoshi Miyawaki[1], *Hisataka Katou*[2], *Masanobu Miura*[3]
[1]*Faculty of Science and Technology, Ryukoku University, Japan,* [2]*Graduate School of Science and Technology, Ryukoku University, Japan,*
[3]*Faculty of Science and Technology, Ryukoku University, Japan*
Poster

ABSTRACT

Proficiency estimations for piano performances were proposed on the basis of a spline curve that represented the tendency of current performance. The tasks used in previous studies were, however, restricted to the scale of one octave or CZERNY40 etude. Their lack of flexibility is therefore a problem since piano pieces played with both hands were not dealt with. In this study, the authors try to estimate the proficiency for the Burgmueller's 25 etude "La candeur". The proposed proficiency estimation method introduces novel evaluation parameters considering a piano piece that is played with both hands and estimates the proficiency by using proposed parameters together with the previous parameters. The proposed method is evaluated using the following conditions: "AP(All Parameter)", "PP(Previous Parameter), "HP(Hand Parameter only)", and "CP(Chord Parameter only)". Results show that adjusted R^2 are .73, .56, .49 and .51 in the case of AP, PP, HP, and CP conditions, respectively. Adjusted R^2 in previous studies were .45(n=336) and .57(n=200) in the case of a scale of one octave and CZERNY40 etude, respectively. The AP condition is thus confirmed to be the most effective condition and to be much higher than in previous studies. Therefore, a proficiency estimation method is achieved for piano pieces played with both hands.

1. INTRODUCTION

Proficiency estimations for piano performances were proposed in previous studies on the basis of a spline curve representing the tendency of current performance [1-4]. The tasks used in previous studies were, however, restricted to the scale of one octave or CZERNY40 etude, which is hard to think of as a performance task for artistic expression. Their lack of flexibility is therefore a problem since piano pieces played with both hands were not dealt with. Here, we employ a short etude extracted from "La candeur" in Burgmueller 25 etude, as a performance task played with both hands. The "La candeur" is thought to be a more practical piece than a scale of one octave or CZERNY40 etude in terms of musicality. In addition, we may solve the problem of proficiency estimation for piano performances in terms of applicability by using "La candeur" as a performance task. The performance task is shown in figure 1, which shows fingering numbers of the right hand from 1 to 5, and those of the left hand from i to v in dashed-line boxes. The proposed proficiency estimation method introduces novel evaluation parameters considering piano pieces played specifically with both hands and estimates the proficiency by using proposed parameters together with the previous parameters. We describe the whole tendency by the spline method as previous studies [1-4] did.

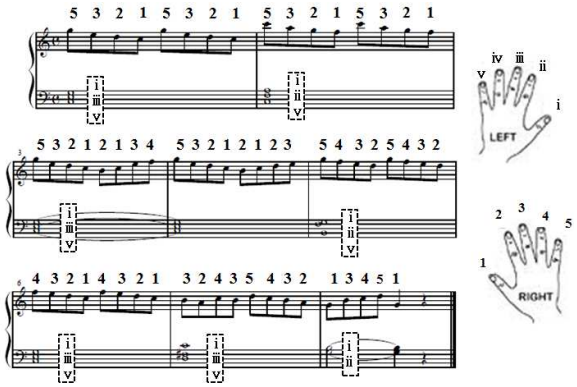


Figure 1: Short etude extracted from La candeur

2. RECORDING EXPERIMENT AND EVALUATION EXPERIMENTS

2.1. Record experiment

This study tries to realize proficiency estimations for "La candeur". We employed 16 people, two expert pianists and 14 students of our university, as subjects of our recording experiment. They were asked to play the performance task in figure 1, and we recorded MIDI and audio of their performance. In addition, we played only eight beats back to subjects as pre-counts. The tempo played back was 120 bpm, and pianists began to play it by following the eight precounts. Pianists were asked to play them ten times for the task. As a result, the performance data did not include any performance errors of omission, insertion, or substitution of notes. The performance data totaled 160 patterns. Piano playing experience of the pianists in the recording experiment was not uniform. Specifically, the two experts had been playing for 35 and 36 years. In addition, students' experience ranged from 1 to 5 years for five students, 6 to 10 years for four students, and more than 11 years for the other five students.

2.2. Evaluation experiment

Using the 160 patterns, we conducted a subjective evaluation, by employing three expert pianists as evaluators. Using the recording audio data of the 160 patterns, we asked the evaluators to rate the pianists' proficiency from one to ten (10 being perfect). To investigate the commonality of their results, we then evaluated correlation coefficients among the three evaluators. We found out that the correlation coefficients ranged from .61 to .67 (avg.= .63, n=160). In a relevant study of a piano performance, correlation coefficient of the subjective evaluation scores of four evaluators ranged from .21 to .56 (avg.=.38, n=102) in Repp [5]. The present study provided higher correlation coefficients than Repp's because correlation coefficients were higher. From here, we employed the normalized score provided by three evaluators.

3. PROFICIENCY ESTIMATION

3.1. SPLINE CURVE

The spline curve was introduced as a standard to calculate the parameter of current performance. To draw a spline curve, representative points need to be passed by the curve. On the plane where the ordinate represents the deviation and abscissa represents note ID, the deviations are modeled as a spline curve and deviations from standards such as metromic line or so. To obtain this curve, we need to classify the notes into several clusters. In previous studies [1-4], notes were automatically clustered on the basis of the label of turning or crossing of fingering. The present study introduces novel criteria for "jumping", "shortening", and "cross-turning" as the standards. Accordingly, the criteria to classify the notes into note clusters are "crossing", "turning", "jumping", "shortening", and "cross-turning", meaning that if current notes satisfy all criteria, we allocate a boundary of clusters on the notes. The center in each cluster is then regarded as its representative point. The five standards of division of notes are shown in figure 2.

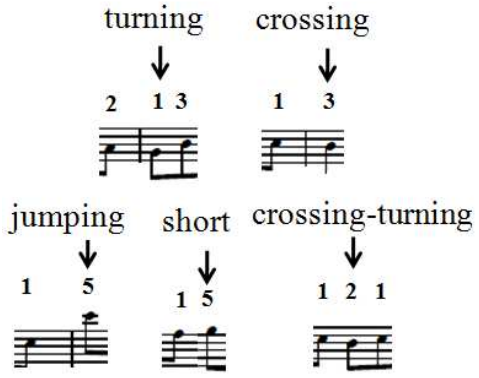


Figure 2: Five standards of division of notes

3.2. Evaluation parameter

This study proposes novel evaluation parameters considering both hands and estimates the proficiency by using the proposed parameters along with the previous parameters. We introduce the following evaluation parameters "keystroke lag between right and left hands" (LagLR) and "keystroke lag composed notes of a chord" (LagCN).

3.2.1. Previous parameters

The evaluation parameter shared with previous methods is referred to as $P^{o}_{wi\gamma}$, where o means previous method, w represents either right hand or left hand, i represent four elements, and _ represent parameters ID. Here, the five evaluation parameters are calculated with both the right hand (w=r) and left hand (w=l) and four elements: onset time (i=t), velocity (i=v), duration (i=d), and instantaneous tempo (i=s). Five evaluation parameters are "standard deviation of subtracting standard from recording", "rms deviation of subtracting standard from recording from spline curve", "range of spline curve", "difference in spline curve between adjacent notes", and "sum of spline curve from standard". The evaluation parameter in the case of the right hand is shown in (3.1) to (3.5), where q represents noteID, $x_{iq}x_{iq}$ value of each element in the piano performance, $\hat{x}_{iq}$ value of each element in the spline curve, $\bar{x}_i$ average value of each element in piano performance, and n the number of input notes. The evaluation parameter is summarized in figure 3, where ordinate shows times and abscissa shows deviation from an ideal score.

・standard deviation of subtracting standard from recording

$$P^{o}_{wi0} = \sqrt{\frac{\sum\left(x_{iq} - \bar{x}_i\right)^2}{n-1}} \qquad (3.1)$$

・rms deviation of subtracting standard from recorded from spline curve

$$P^{o}_{wi1} = \sqrt{\sum_{k=1}^{n}\left|x_{iq} - \hat{x}_{iq}\right|^2} \qquad (3.2)$$

・range of spline curve

$$P^{o}_{wi2} = \max\left(\hat{x}_{iq}\right) - \min\left(\hat{x}_{iq}\right) \qquad (3.3)$$

・difference in spline curve between adjacent notes

$$P^{o}_{wi3} = \sqrt{\sum_{k=2}^{n}\left(\hat{x}_{iq} - \hat{x}_{iq-1}\right)^2} \qquad (3.4)$$

・sum of spline curve from standard

$$P^{o}_{wi4} = \sum_{k=1}^{n}\hat{x}_{iq} \qquad (3.5)$$

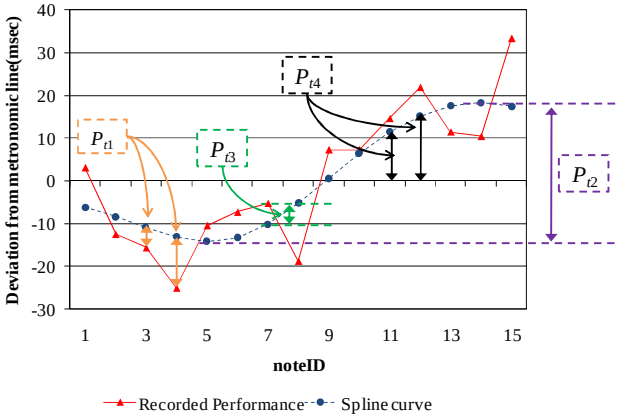


Figure 3: Summary of the evaluation parameter

3.2.2. Keystroke lag between right and left hands

Evaluation parameters for onset time and velocity used in the LagLR are "standard deviation of LagLR", "rms deviation of LagLR", and "range of LagLR". We calculate LagLR by using three different onset times. "La candeur" chords played with the left hand have consistent timing in sheet music. Therefore, we evaluate three differences between a note of the right hand and each note of the left hand. The differences are labeled as RL1, RL2, and RL3, where the numbers are the index of prior (1) to posterior (3) of played timing. Figure 4 shows an example of the calculation method in LagLR, where ordinate shows note height, and abscissa shows time. In addition, L1, L2, and L3 show onset times of

the first, second, and third keying notes in the chord, respectively, R shows onset time of the right hand, and h_1, h_2, and h_3 show RL1, RL2, and RL3, respectively.

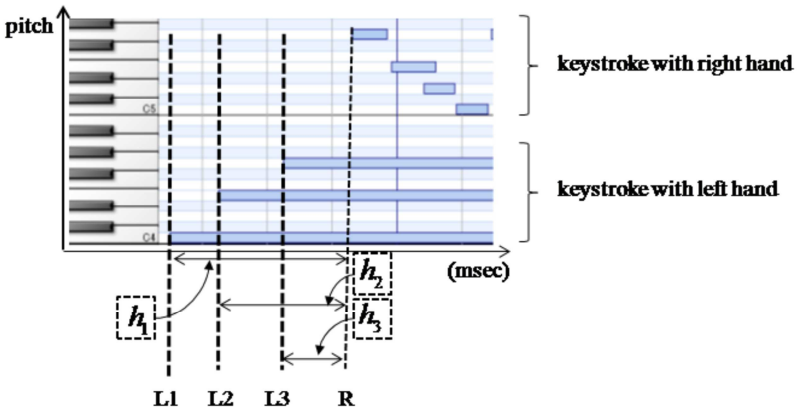


Figure 4: Example of the calculation method of keystroke lag between right and left hands

3.2.3. Keystroke lag of notes in a chord

Evaluation parameters for onset time, duration and velocity used in the LagCN are "standard deviation of LagCN", "rms deviation of LagCN", and "range of LagCN". We use chords composed of three notes. We calculate LagCN by using two different onset times. In the case of "La candeur", notes in chords are simply played simultaneously in sheet music. Therefore, we evaluate the two differences between each note of the left hand. The gaps are labeled as L12 and L13, where the numbers are index of prior (1) to posterior (3) of played timing. They are the differences between the first and the second keying notes in the chord (L12) and those for 1st to 3rd (L13). Figure 5 shows an example of the calculation of LagCN, where ordinate shows note height, and abscissa shows time. In addition, L1, L2, and L3 show onset times of the first, second, and third keying notes in the chord, respectively, and c_{12} and c_{13} show L12 and L13, respectively.

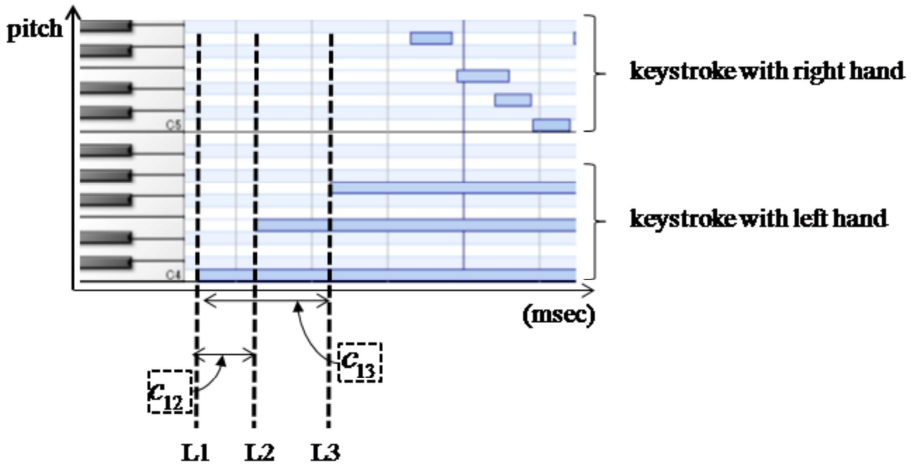


Figure 5: Example of the calculation method in keystroke lag of notes of a chord

3.2.4. Relationship between evaluation parameter and proficiency score

We investigate to what degree 59 evaluation parameters correlate with the proficiency evaluation. Specifically, we calculate the correlation coefficient between subjective evaluation score and the evaluation parameter. Results are shown in figure 6.

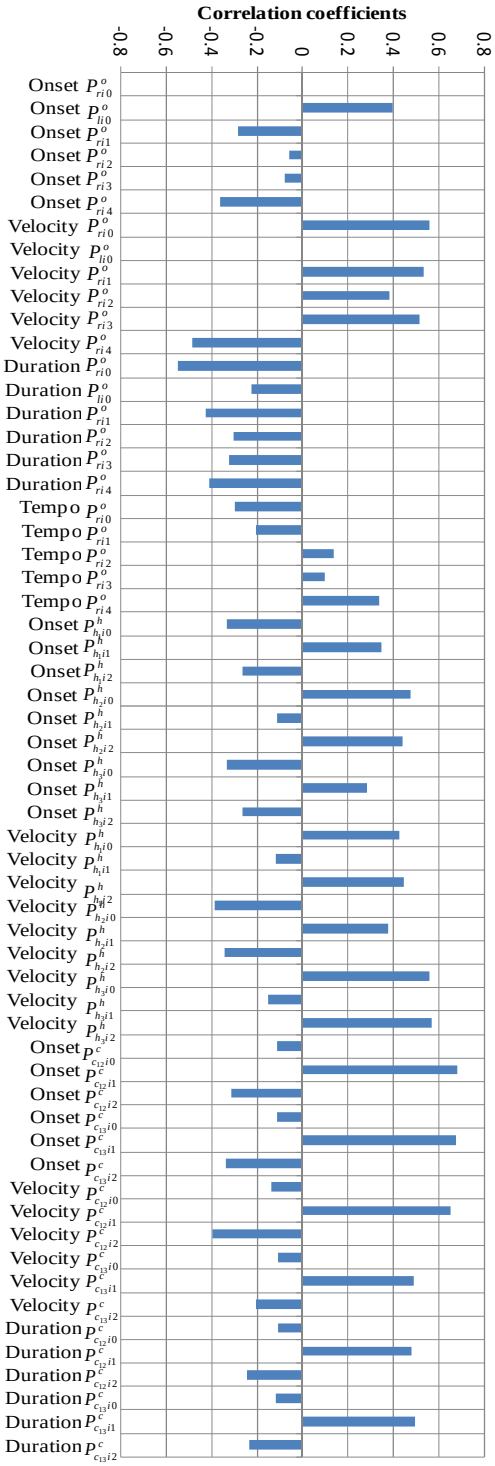


Figure 6: Correlation coefficient between subjective evaluation score and evaluation parameter

3.3. Proficiency estimation

In the first estimation, the current performance's tendency curves for each deviation are calculated by using spline curves for onset time, velocity, duration, and instantaneous tempo. Deviations for onset time, velocity, duration, and tempo are then obtained. Finally, 59 parameters in total are obtained. By decreasing the dimension using principal component analysis, 16 principle components are obtained. The nearest five samples are then chosen out of other samples in neighborhoods. The average score of the five samples is considered as the score of proficiency for input performance. Coefficients of determination are evaluated between the estimated scores given by the proposed method to the evaluated scores given by expert pianists. Proficiency estimation is summarized in figure 7.

4. RESULTS

4.1. Cross validation

Here, two types of cross validations ("leave-one-out cross validation" and "16-fold cross validation") were employed to evaluate the effectiveness of the proficiency estimation method.

- leave-one-out cross validation

Of the 160 performance datasets provided by recording ten performances of each subject, 159 are dealt with as the training data and the remaining one is evaluated. The step is repeated 160 times.

- 16-fold Cross Validation

Of the 160 performance datasets provided by recording ten performances of each subject, 150 are dealt with as the training data and the remaining 10, which are from one pianist, are evaluated. The step is repeated 16 times.

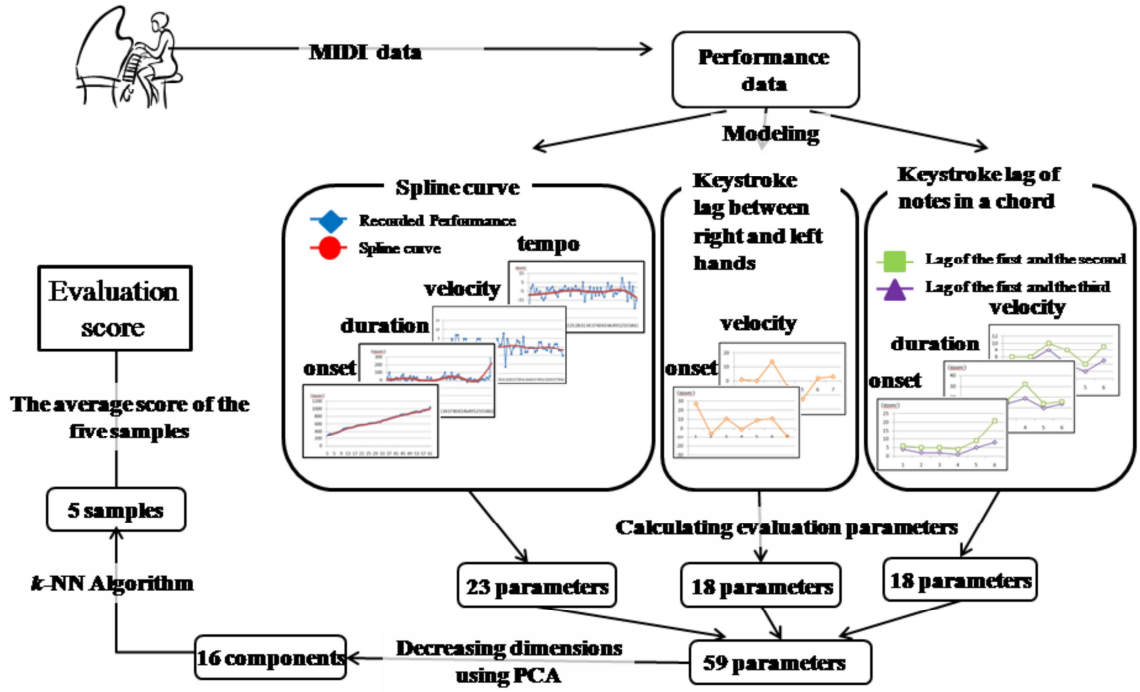


Figure 7: Summary of proficiency estimation

4.2. Proficiency estimation by proposed method

The proposed method is evaluated using the following conditions: "AP (All Parameter)", "PP (Previous Parameter), "HP (Hand Parameter only)", and "CP (Chord Parameter only)". Also, we compare the precision of estimated scores for two methods of cross validation. We calculate correlation coefficients and adjusted R^2. The correlation coefficients and adjusted R^2 by leave-one-out cross validation are shown in figure 8, and those for 16-fold cross validation are shown in figure 9.

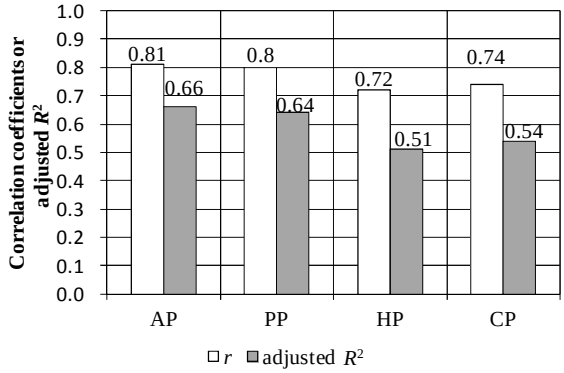


Figure 8: Correlation coefficients and adjusted R^2 of proficiency score and the evaluation score in the leave-one-out cross validation

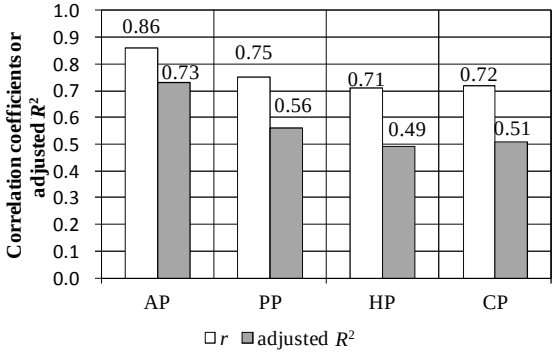


Figure 9: Correlation coefficients and adjusted R^2 of proficiency score and the evaluation score in 16-fold cross validation

5. DISCUSSION

From figures 8 and 9, the results of AP were the highest in correlation coefficients and adjusted R^2 . The second highest was PP, followed by HP and CP. Therefore, the AP was confirmed to be the most effective evaluation parameter set for the proficiency estimation. In addition, CP was confirmed to be an effective evaluation parameter for the proficiency method because it was slightly higher than HP in all cases of correlation coefficients and adjust R^2.

However, no salient difference was confirmed between HP and CP. We then compare the previous and proposed parameters (HP and CP) and found that previous parameters (PP) were higher than HP and CP in correlation coefficients and adjusted R^2 from figures 8 and 9. However, AP was higher than PP in the value of correlation coefficients and adjusted R^2. Thus, both CP and HP surely contributed to proficiency estimation, and AP was confirmed to be an effective evaluation parameter set for our proficiency method.

An evaluation parameter set that contains previous parameters (PP) and proposed parameters (HP and CP) was confirmed to be effective. In addition, results of leave-one-out cross validation are somewhat unreliable because some performance data of the pianist for current input data are included in supervised data, and the test is essentially closed rather than open. On the other hand, the performance data of the pianist for current data are not included as supervised data in the 16-fold cross validation. Therefore, these data could maintain their objectivity and be more reliable than leave-one-out cross validation. The results of the proficiency estimation using 16-fold cross validation were r=.86(n=160) and adjusted R^2=.73(n=160) on the AP. We compared the correlation coefficients and adjusted R^2 between proficiency scores and subjective evaluation scores and found that this study provided a higher proficiency estimation score than previous studies [1,6], as shown in table 1. Thus, our results for correlation coefficients and adjusted R^2 in this study were higher than those in previous studies. As a result, higher precision was provided in our case, and the effectiveness of the proposed method was confirmed.

	Miura (2010)	Nonogaki (2012)	This study
Performance task	Scale of one octave	CZERNY 40 etude	La candeur
r	.65 (n=336)	.76 (n=200)	.86 (n=160)
adjusted R^2	0.45	0.57	0.73

Table 1: Comparison of performances among relevant studies

6. CONCLUSIONS

This study proposed a proficiency estimation method for the Burgmueller 25 etude “La candeur” played with both hands. In addition, we obtained proficiency scores provided by expert pianists to use in proficiency estimation. As a result, the performance data did not include any performance errors of omission, insertion, or substitution of notes. In total, 160 performance datasets were obtained. The evaluators listened to these performances and evaluated each from one to ten (10 being perfect). In addition, correlation coefficients and adjusted R^2 in this study were compared and showed higher proficiency estimation than in previous studies [1,6]. As a result, it was thought that a certain precision was provided, and the effectiveness of the proposed method was confirmed. Future plans are to examine other parameters that grasp the characteristic of the performance more exactly, establish another estimation method that is more precise than the present method, and apply it for various musical pieces.

7. REFERENCES

Miura, et al., *“Automatic evaluation of performance proficiency of one-octave major scales using the piano”, Journal of the Acoustical Society of Japan, 66, pp.203-212 (2010, in Japanese)*

Nonogaki, et al., *“Evaluation parameters for proficiency estimation of piano based on tendency of moderate performance”, Proc. of 12th International Conference on Music Perception and Cognition (ICMPC), pp.728-737 (2012)*

Morita et al., *"Evaluation of a scale performance on the piano using spline and regression models", Proc. of International Symposium on Performance Science, pp.77-82 (2009)*

Shimazu et al., *"Method for estimating proficiency in playing acoustic piano", Proc. of The 20th International Congress on Acoustics, 364 (2010)*

Repp, B., “A microcosm of musical expression. III. Contributions of timing and dynamics to the aesthetic impression of pianists’ performances of the initial measures of Chopin’s Etude in E Major”, J. Acoust. Soc. Am.104, pp.1085-1100 (1998)

Nonogaki, A., “Proficiency estimation of piano piece performance based on the artistic deviation”, master thesis, school of science and technology, Ryukoku University of Japan (2012)

[3-16] GESTALT UNIVERSALS OR MUSICAL ENCULTURATION: PERCEIVING BOUNDARIES IN UNFAMILIAR TURKISH MAKAM MUSIC

Z. Funda Yazici[1], Esra Mungan[2], Mustafa Uğur Kaya[3]

[1]Istanbul Technical University, Music Theory and Musicology, Turkey, [2]Boğaziçi University, Psychology, Turkey, [3]Koç University, Psychology, Turkey

Poster

Gestalt theory predicts perceptual grouping of elements based on their proximity to each other, their similarity or nonsimilarity with each other, as well as other principles such as good continuation and closure. As such, Gestalt theory has a universal stance towards any kind of perceptual grouping, including auditory grouping, regardless of the type of material or type of listener groups used. Lerdahl and Jackendoff (1983), for instance, used Gestalt laws to develop a musical model of grouping for Western hierarchical music. This model has been put to test extensively within the Western musical context and studies have indeed shown considerable agreement between perceptual segmentations of musically trained and nontrained listeners both for classical music pieces (e.g., Deliège, Mélen, Stammers, Cross, 1996) as well as for 20th century modern music (e.g., Deliège & El Ahmadi, 1990). Our goal was to see whether Gestalt laws of perceptual organization would also predict segmentation

in non-Western music. In our study, musically trained and untrained Turkish listeners, and Western listeners segmented unfamiliar Turkish makam music. Participants were asked to put a mark whenever they subjectively perceived a boundary after one pure listening and two segmentation practice trials for each of eight ca. 1-min-long 19th century Turkish makam tunes. All tunes were recorded on MIDI carefully retaining their microtonal structure. Results showed that Turkish listeners showed high within- as well as between-group agreement in their segmentations. More interestingly, Western listeners completely unfamiliar with makam music culture, also showed considerable within- as well as between-agreement as far as major segmentation points were concerned. Participants' perceptual groupings were also compared with expert segmentations and basic segmentations predicted by Gestalt-based laws. The effects of musical training and musical enculturation were astonishingly low. Overall, results suggest substantial universality in perceptual grouping as predicted by the two most basic Gestalt laws, proximity and similarity.

[3-17] CAN INDIVIDUALS, BOTH TRAINED AND UNTRAINED, BE MADE AWARE OF THEIR INVOLUNTARY RESPONSES TO TONE RELATEDNESS AND, IF SO, WOULD IT MAKE ANY DIFFERENCE?

Nancy Garniez[1]
[1]*Mannes College, The New School for Music, USA*
Poster

The study aimed to determine whether individuals could be made aware of their involuntary experience of tone relatedness, demonstrated by Viktor Zuckerkandl (Sound and Symbol) to be integral to the musical life of persons with or without training; and, if so, whether it would make any difference. Seventy-two individuals, including professional musicians, amateurs, listeners, and children, were seen for three weekly sessions of 1 ¼ hours. Session I: Each was asked to identify a tone that had particular meaning for them, then to choose out of hundreds of colored pencils a color to match that tone. Proceeding then to relate that tone/color either to a scale, or directly to the tones in a specific composition of their own choosing, the individual drew the tones on a grid, as if depicting vibrations. With the horizontal axis of the grid corresponding to time measured objectively, the vertical axis registered degrees of rise and fall as perceived by the inner ear, i.e., subjectively, often deviating significantly from standard notation. Continuing to depict the composition the individual worked in silence, relating to the score without relating physically to an instrument or actual sound. Session II: A continuation of the work. By Session III 67% of individuals already registered a clear response, generally manifested in improved reading; and better, more confident coordination in instrumental or compositional performance. Of this group 13% presented acute physical or psychiatric symptoms; with the exception of only one individual in this sub-group, the results were dramatically restorative. For 30% of participants the process seemed to have no evident relevance while, for 3%, the first two sessions evoked such psychological pain as to preclude completing the study. Representative samples of work from all three groups will be shown. This research has direct implications about relating visual to auditory experience. Whereas failing to account for the difference between fully resonating sound and the discrete symbols of standard notation risks alienating the reader from innate musical sense, a system that translates auditory experience into visual terms may restore the connection.

[3-18] FLEXIBLE MODULATION IN SPATIAL REPRESENTATION OF THE PIANO KEYBOARD IN TRAINED PIANISTS

Chie Ohsawa[1,2], *Satoshi Obata*[3], *Takeshi Hirano*[4], *Minoru Tsuzaki*[2], *Hiroshi Kinoshita*[3]
[1]*Japan Society for the Promotion of Science, Japan,* [2]*Faculty of Music, Kyoto City University of Arts, Japan,* [3]*Graduate School of Medicine, Osaka University, Japan,* [4]*Faculty of Management, Osaka Seikei University, Japan*
Poster

1. Background: Trained pianists had relatively inaccurate spatial memory of key positions outside regions to the keyboard centre (Ohsawa et al., 2012, 2013). This finding was surprising because in real performance, pianists rarely make mistakes in locating the fingers, even in rapid octave leaps.

2. Aims: We conjectured that pianists often move and adapt their central region of the keyboard memory to the playing area. To test this hypothesis, we examined accuracy of key position detection in the remote keyboard areas after giving an attentional shift of keyboard area.

3. Method: Fourteen pianists (with at least 15 yrs of piano training) participated in the experiment. Each subject faced a table on which a flat sheet of either the whole 88-key keyboard or only a single reference key (C4 or C7) was printed. The subject located his/her right index fingertip on the reference key first, and then moved to the target key (C3, E3, A4, C5, C6 and E6 with the reference of C4, and A4, C5, C6, E6, A7 and C8 with the reference of C7). These were shown by a score. These target notes were randomly presented. For each of the reference and target notes, the participants touched the designated key for 3 sec, followed by a 3-sec inter-note rest. Kinematics of the fingertip was recorded using 3D motion capture system for 10 trials for each target key.

4. Esults: For the target notes of A4, C5, C6 and E6, the pointing accuracy (absolute errors of horizontal distance from the target key center) significantly differed between the C4 and C7 reference conditions. On the other hand, when the distance of target key position from the reference key was matched between the two conditions, the condition difference disappeared.

5. Conclusions: We concluded that pianists flexibly modulate their spatial representation of the keyboard by locating attention to the area for upcoming working area. This ability seems to be developed in sighted pianists to fully use the 88 keys in order to minimize the risk of touching an adjacent key.

[3-19]Evidence that Phrase-Level Tempo Variation May be Represented Using a Limited Dictionary

Shengchen Li[1], *Dawn Black*[1], *Elaine Chew*[1], *Mark D. Plumbley*[1]
[1]*Queen Mary University of London, United Kingdom*
Poster

ABSTRACT

Phrases are common musical units akin to that in speech and text. In music performance, performers often change the way they vary the tempo from one phrase to the next in order to choreograph patterns of repetition and contrast. This activity is commonly referred to as *expressive music performance*. Despite its importance, expressive performance is still poorly understood. No formal models exist that would explain, or at least

quantify and characterise, aspects of commonalities and differences in performance style. In this paper we present such a model for tempo variation between phrases in a performance. We demonstrate that the model provides a good fit with a performance database of 25 pieces and that perceptually important information is not lost through the modelling process.

1. INTRODUCTION

We posit that, like melody perception, humans do not apprehend individual tempo values; instead, we grasp tempo patterns at the figural and phrase level. Furthermore, we propose that the vocabulary of acceptable tempo patterns is a subset of all possible variations, and might be represented by a limited dictionary of tempo patterns.

As a precursor to any research on tempo pattern representation, the natural question arises: Are human listeners aware of tempo pattern variation? And, do we prefer variations in tempo patterns? Indeed, if the humans can discern when tempo variation patterns are employed in performance, then any model of expressive performance, and by extension any model of music (because music is communicated through performance,) must consider tempo variation. Furthermore, we are concerned with whether human listeners are aware of degrees (magnitudes) of tempo variation through a performance. We design experiments to validate a perceptually weighted measure of tempo variation, the Tempo Variegation Coefficient (TVC).

A goal of this research is to provide supporting evidence for the existence of a dictionary of phrase-level tempo patterns. An analysis and grouping of actual tempo changes allows us to extract a dictionary of common tempo patterns. In this paper we process tempo information in two stages: first we segment the music into two-bar phrases and create tempo curves of a fixed score length; next, these two-bar length curves are clustered using a Gaussian Mixture Model (GMM) to create a dictionary with *n* components. Using the Log Likelihood measure to quantify how well the (GMM) model predicts tempo patterns, we show that the model is a good fit for actual performances both statistically and perceptually.

In order to quantify the model's fit with listener perception we propose a new measure, the Tempo Variegation Coefficient (TVC), that quantifies the inter-phrase variation in tempo. Next, we show a direct correlation between the magnitude of the TVC and listeners' perception of tempo variation for the recorded performances. We show that the relative degree of tempo pattern variation predicted by the TVC using the clustered phrase-level tempo data is closer to listener perception than the TVC computed using the original phrase-level tempo data. The dictionary of tempo patterns is thus shown to successfully model expressive performances both statistically and perceptually.

1.1. Related Work

It has been proposed by Repp (Repp, 1999) that 'actual performances are only a subset of possible performances.' In the context of tempo variation, we can therefore surmise that performers choose from a limited number of expressive tempo choices or tempo patterns. This general set or 'dictionary' of tempo patterns, if found, would constitute a model for musical expression (in tempo) and provide insight into the communication and understanding of musical expression.

Creating a dictionary invariably involves an element of aggregation, grouping, or generalisation. Other researchers have sought to find patterns in tempo information in order to better characterize performances. In (Repp, 1995a) Repp used Principal Component Analysis to determine the generating bases for tempo trajectories across entire pieces. He demonstrated that performances by piano students have only one primary PCA component, compared to professional pianists whose performances result from more components. Repp later improved upon this method by applying Varimax Rotation (VR) to the principle components (Repp, 1999). Spiro (Sprio, Gold, & Rink, 2010) used Self Organizing Maps to classify bar-length tempo and loudness patterns into four types to characterize performances. The usage distribution of the bar types was shown to differ among performers.

Patterns found in performance data have also been used for synthesizing expressive performances. The DISTALL system (Widmer & Tobudic, 2003) compares phrases in a piece of music, and formulates an expressive timing strategy based on music structure. In (Widmer, Flossmann, & Grachten, 2010), Widmer et al. used a Bayesian model, YQX, to synthesize expressive timing and dynamic variation. For a more comprehensive review of the state-of-the-art, see (Kirke & Miranda, 2013).

Much remains to be learned about human perception of tempo. Previous research includes studies establishing the Just Noticeable Difference (JND) for tempo, studies on the human ability to recall tempo, and studies on other aspects of tempo perception. Levitin (Levitin, 1996) found that performers are more adept at memorising absolute tempo than relative tempo. Dahl and Granqvist (Dahl & Granqvist, 2003) proposed a 2%, and Levitin (Levitin, 1996) a 3.5-11% JND threshold (Levitin, 1996). In the present study, we are concerned with not only the discernment of JND for static (or stable) tempi, but of tempo patterns.

In (Cambouropoulos, Dixon, Goebl, & Widmer, 2001) the authors demonstrate human preference for smooth tempo changes. This evidence for the perceptual relevance of tempo smoothing is significant because it implies that our use of a dictionary of general tempo curves may actually improve a model's correlation with human perception. In a study on relative modulation depth (a measure of tempo variation), Repp (Repp, 1995) shows that listeners may judge the quality of performance based on the degree of tempo changes.

This paper is organized as follows: first, we describe the construction of our performance database. We then describe the grouping or clustering of our performance into candidate dictionaries. We assess the appropriateness of these dictionaries by measuring their ability to model an actual performance. Their accuracy is measured in two ways: quantitatively using the maximum likelihood; and, qualitative through comparison with listeners' perception of tempo variation.

2. PERFOMANCE DATABASE

In order to analyse and model tempo variation across phrases, a tempo dataset is required. For this work we have chosen Mily Balakirev's *Islamey*, Op 18 (Balakirev, 1902)**.** This piece has two advantages as an exemplar piece when considering tempo variations across phrases. First, the piece exhibits a straightforward hierarchical phrase structure. Basic phrases are two bars long and every two or four two-bar phrases combines to form a higher-level phrase. This allows for unambiguous analysis at the phrase level. Secondly, it contains numerous repetitions of the main theme. The existence of repetitions can be viewed as introducing an element of periodicity, and hence predictability or invariance, into the performance, which assists with the formation of a small, well defined dictionary.

The phrase-structure of the first 12 phrases of Islamey is shown in Figure 1. The dark and light shading delineate two-bar (8-beat) phrases. The labels 'A', 'B' and 'C' indicate the note content type, highlighting the repetitions in the structure.

A database comprising 25 performances of Islamey performed by a range of well-known performers was created and all the beats in the first 84 bars (336 beats or 42 phrases per performance) were hand annotated using Sonic Visualiser (www.sonicvisualiser.org). To track tempo changes, a tempo time series linking all beats was created for each phrase, resulting in an 8-beat tempo curve. Example tempo curves can be seen in Figure 3.

3. HUMAN PERCEPTION OF TEMPO

Models of expression can be used to characterize and quantify commonalities and differences in performance style. However, a model is not useful unless the performance element it emulates influences human perception. The performance element we examine is tempo variation and, in order to support the need for a model of tempo variation, we show that humans are aware of tempo variation and that they can differentiate between degrees of tempo variation.

In order to quantify the degree of tempo pattern variation in a performance, we propose a new measure, the TVC, described in Section 3.3. We use a listening test to show a strong correlation between the magnitude of the TVC and human perception of tempo variation.

3.1. Are humans aware of tempo variation?

We designed a listening test to answer the question of whether human listeners are aware of tempo variation. Eighteen listeners (seven with no musical training, eight having between 1 and 10 years of training, and three with >10 years of training) listened to 9 pairs of MIDI files, each with an identical melody: one having tempo variation and one having no tempo variation. Listeners were asked to indicate which MIDI sample included tempo variations. They were also asked to indicate which sample they preferred. All MIDI was synthesized from the score of *Islamey* (our case study) and tempo variations were based on simplification tempo trajectories from our performance database, and hence could be considered natural but not a fully expressive performance. The number of tempo patterns employed ranged were: one (the average of all tempo curves); two (the centroids of a two-component GMM trained on our performance database); and, three (the centroids of a three-component GMM).

The survey results are presented in Tables 1 and 2. Listener preference is mapped to a scale of 1-5 (‘1’ indicating strong, and ‘2’ indicating slight, preference for the MIDI without tempo variation; ‘3’ no preference; and, ‘4’ indicating slight, and ‘5’ indicating strong, preference for the MIDI with tempo variation).

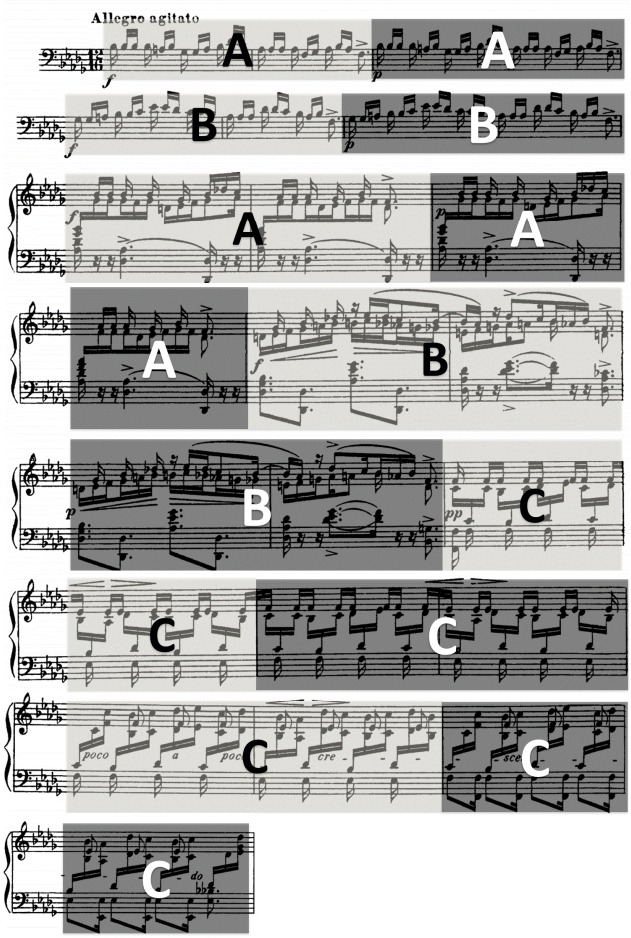


Figure 1: Structure of Balakirev’s *Islamey*.

Table 1: Results of the tempo variation perception listening test.

Number of tempo patterns	1	2	3
Listeners choosing MIDI with tempo patterns (%)	**72**	**83**	**94**
Tempo pattern preference score (variance)	**3.44 (1.08)**	**3.11 (1.86)**	**3.06 (2.03)**

These results indicate that humans are aware of tempo change and, on the examples we used, people showed a slight preference for music that exhibited variations in tempo.

3.2. Human response to degrees of tempo variation

Having demonstrated that humans are cognizant of, and have a slight preference for, variation in tempo, the second question can be addressed: can humans differentiate between degrees of tempo variation. When played two samples of the same melody, one of which can be shown mathematically to exhibit greater variation in tempo than the other, are listeners able to identify the sample with the greatest tempo variation?

Six listeners, all of whom were in training to become professional pianists, were asked to listen to sets of three samples from our performance database, and to sort the samples in order of perceived variation in tempo. Performers chosen from our database were labeled P2, P3, P4, P5, P7, P12, P15, and P17. These performers were chosen because preliminary listening tests showed that these performers differed most markedly in their expressive strategies. Six sets of three performances each were randomly selected and listeners were asked to order the three performances from one exhibiting the most tempo variation to the one exhibiting the least tempo variation. Results are presented in the first column of Table 2. Table 2 shows that, when presented with performers P2, P7, and P15 (row 1, column 1), listeners ordered them in the following manner: P15 (most tempo variation), P7 (less tempo variation than P15 but more than P2), and P2 (least tempo variation). This suggests that listeners are aware of degrees of tempo variation. We now need a measure of tempo variation that shows a strong correlation to listener perception. We propose a novel measure: the Tempo Variegation Coefficient.

Listener order	**TVC1 (tempo curves)**	**TVC2 (cluster centroids)**
P2 > P7 > P15	**P2** > P15 > P7	P2 > P7 > **P15**
P3 > P4 > P25	P4 > P3 > **P25**	**P4 > P3 > P25**
P12 > P5 > P17	**P12 > P5 > P17**	**P12 > P5 > P17**
P2 > P5 > P17	**P2 > P5 > P17**	**P2 > P5 > P17**
P12 > P7 > P25	**P12 > P7 > P25**	**P12 > P7 > P25**
P3 > P4 > P15	P4 > P3 > **P15**	**P4 > P3 > P15**

Table 2: Results from the degree of tempo variation listening test and correspondence to the Tempo Variegation Coefficient. Bold text indicates agreement between listeners and the TVC.

3.3. Tempo Variegation Coefficient (TVC): A measure of perceived tempo variation

We can compare tempo variation at the inter- or intra-phrase level. Intra-phrase tempo variation measures tempo variation between individual beats in a phrase. Inter-phrase tempo variation measures tempo variation between adjacent phrases. In order to show different degrees of tempo variation, we introduce the term 'tempo variegation' to denote the distance between tempi in corresponding beats in different phrases. For the purpose of creating a dictionary of phrase-level tempo patterns the inter-phrase tempo variegation is measured.

In our case study *Islamey*, a phrase is 8-beats long. A performer will play each phrase with differing patterns of tempo variation in order to communicate their expressive strategy. The tempo curves employed by a performer can be considered a set of phrase level tempo vectors, $\vec{A} = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$, where $\vec{x}_i$ represents the tempo curve employed in the *i*-th phrase. Following clustering, suppose the corresponding cluster centroids of the tempo vectors are given by $\bar{A} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$. The TVC effectively measures the average Euclidean distance between corresponding tempi in different phrases. We introduce two variants on the TVC. TVC1 is calculated using the original tempo curves:

$$\mathrm{TVC1} = \frac{2}{n(n-1)} \sum_{\vec{x}_i, \vec{x}_j \in \vec{A}} \left\| \vec{x}_i - \vec{x}_j \right\|_2 , \qquad (1)$$

where *n* is the total number of phrases. TVC2 is computed using the cluster centroids of each tempo curve:

$$\mathrm{TVC2} = \frac{2}{n(n-1)} \sum_{\bar{x}_i, \bar{x}_j \in \vec{A}} \left\| \bar{x}_i - \bar{x}_j \right\|_2 . \qquad (2)$$

The TVC's can be considered as measures of the tempo variegation within a performance, where a higher TVC value indicates a greater degree of tempo variegation. Figure 2 shows the TVC2 value for each performer in our database. The greater the magnitude of the TVC, the more tempo variegation that performer is considered to exhibit. Columns 2 and 3 in Table 2 order each set of three performances according to the magnitude of their TVC1 and TVC2 values, respectively, i.e., their tempo variegation according to the original tempo curves and the clustered tempo curve centroids, respectively. We can see that the TVC1 rankings concur with listener judgments in three out of the six triplet performances; the TVC2 rankings concur with five out of the six triplet performances. This demonstrates a strong correlation between the magnitude of TVC2 and human perception. TVC2's predictions show better correspondence to listener judgments than that of TVC1, which suggests that listeners may impute some degree of tempo smoothing when perceiving tempo change.

4. DICTIONARY CREATION

4.1. Tempo Curve Clustering

Having shown that humans are aware of tempo change and degrees of tempo change, we have justified the need for a perceptually relevant model of tempo change as a component of an overall model of expressive performance. Having surmised that there exists a limited number of expressive tempo choices or tempo patterns for performers to choose from, we now proceed to extract this dictionary from our performance database.

GMMs are commonly used to group or cluster data. In this work, they are used to cluster our database of tempo curves. We first normalize our tempo curves to remove performer bias due to differing overall speed. The scaled tempo curves, ranging between 0 and 1, is computed using the following formula:

$$S_j^{norm} = \frac{s_j - \min(\boldsymbol{r}_p)}{\max(\boldsymbol{r}_p) - \min(\boldsymbol{r}_p)}, \text{for j} = i_p, \dots, i_p + k, \quad (3)$$

where $\boldsymbol{r}_p^{norm} = [s_{i_p}^{norm}, \dots, s_{i_p+k}^{norm}]$.

In order to use the normalized tempo curves to train our GMM we must first determine how many clusters are required. An *n* cluster GMM will result in a dictionary with *n* entries. The aim of the dictionary should be to provide the best performance model whilst using the fewest entries possible.

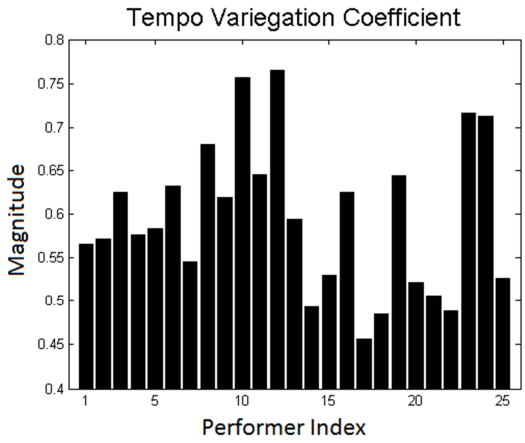


Figure 2: Tempo Variagion Coefficient (TVC2) for each performer in our database.

4.2. Statistical Evaluation of Dictionary Suitability

We can assess our dictionary's suitability by measuring its ability to model expressive performance. Using the leave-one-out method of cross-validation, we train the GMM using 24 performances, then use the resulting dictionary of *n* clusters to predict the tempo variation in the remaining performance. The accuracy with which we can predict the 'left-out' performance is quantified using the maximum-likelihood measure. We vary the number of clusters (Gaussian components) from 1 to 40, and cycle through each combination of training-test data to compute the average likelihood. Figure 3 shows the results of this experiment.

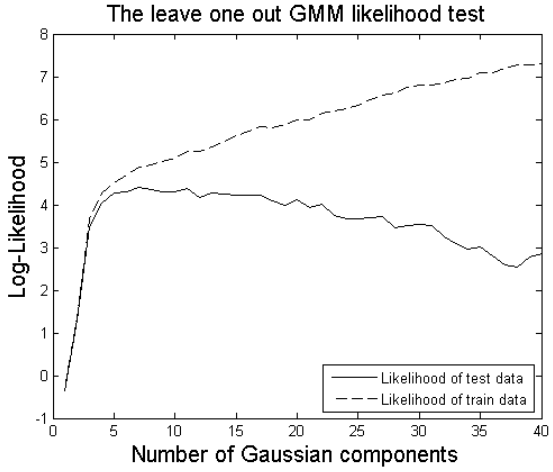


Figure 3: The results of the GMM leave-one-out likelihood test.

Figure 3 shows that the dictionary accuracy when modelling tempo increases sharply when the number of Gaussian components increases from 1 to 5, and remains relatively constant between 5 and 13. Thus, we choose a dictionary size of 8. Figure 4 shows the eight GMM clusters produced with their centroids marked in bold. These centroids form the entries in our dictionary.

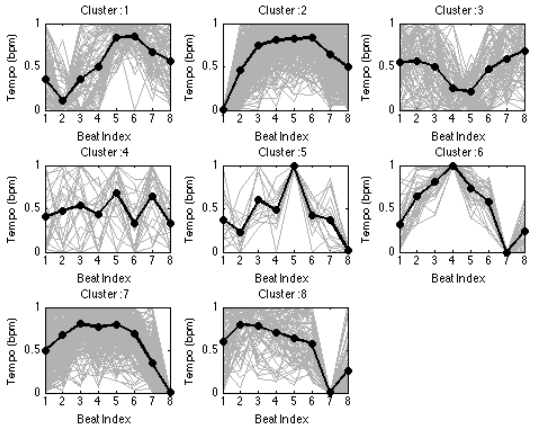


Figure 4: Grouping of tempo curves using a GMM with 8 clusters. Grey lines show individual tempo curves; bold lines depict cluster centroids.

4.3. Perceptual Evaluation of Dictionary Suitability

It is important that any model of expression, consisting as it does of generalizations, preserves perceptually salient information. A model of expressive performance is not useful if, during the creation of that model, information that communicates some significant element of expression to a human listener is removed. We show strong correspondence between listener judgments and TVC2, the Tempo Variegation Coefficient computed using cluster centroids..

Table 2, first discussed in Section 3.2, shows how professional pianists ordered sets of three performances based on perceived tempo variation. Using the TVC to quantify the degree of tempo variation in a performance we repeat the experiment and order the performances according to their TVC1 and TVC2 magnitudes in columns 2 and 3, respectively. Recall that TVC is computed using the actual tempo curves, while TVC2 is calculated using the tempo cluster centroid of each phrase as provided by our dictionary.

The results are convincing. Using our dictionary, five out of the six triplet performances were ordered 'correctly' (in agreement with human listeners). When the 'true' tempo curves are used (not our dictionary), the results were worse. Only three out of the six performance triplets are ordered correctly. This can be considered supporting evidence for the suggestion in (Cambouropoulos, Dixon, Goebl, & Widmer, 2001) that the perception of tempo is better approximated by smoothed tempo.

5. DISCUSSION

We have shown that, by modelling the tempo variation within a performance, we are able to provide a dictionary of general tempo patterns commonly adopted by performers for the piece *Islamey*. This dictionary has been demonstrated to model expressive performance well from both a statistical and perceptual perspective for our database. The existence of the dictionary suggests that performers do indeed constrain themselves to a subset of all possible performance gestures, at least for tempo variation. The clustered results shown in Figure 4 suggests the existence of common performance strategies. For example, there are no curves that show tempo deceleration in the first half of the phrase and acceleration towards the latter half.

6. ACKNOWLEDGEMENT

This research is funded in part by the Chinese Scholarship Council.

7. REFERENCES

Balakirev, M. (1902). Islamey, Op. 18. *D. Rahter.*

Cambouropoulos, E., Dixon, S., Goebl, W., & Widmer, G. (2001). Human Preferences for Tempo Smoothness. Proceedings of the VII Internaional Symposium on Systematic and Comparative Musicology. III International Conference on Cognitive Musicology, *(pp. 18-26).*

CostaP.Lamartine. (1969). Capoeira sem Mestre. Rio de Janeiro: Tecnoprint.

CsordasJThomas. (1993). Somatic Modes of Attention. Cultural Anthropology, 8 (2), 135-156.

Dahl, S., & Granqvist, S. (2003). Looking at Perception of Continuous Tempo Drift - A New Method for Estimatin Internal Drift and Just Noticeable Difference. Proceedings of the Stockholm Music Acoustics Conference, *(pp. 595-598).*

deMoorMaartin. (无日期). Jogo de Capoira Angola. 检索日期: 2014 年 May 月 31 日, 来源: www.wesleyan.edu: http://www.wesleyan.edu/wsa/capoeira/articlejogo.html

DowneyGreg. (2005). Learning Capoeira: Lessons in Cunning from an Afro-Brazilian Art. Oxford: Oxford University Press.

DowneyGreg. (2002). Listening to Capoeira: Phenomenology, Embodiment, and the Materiality of Music. Ethnomusicology, 46 (3), 487-509.

DowneyGreg. (2008). Scaffolding imitation in capoeira: physical education and enculturation in an Afro-Brazilian art. American Anthropologist, 110 (2), 204-213.

GratierMaya. (2008). Grounding in musical interaction: evidence from jazz performances. Musicae Scientiae, 12, 71-110.

HudakPaul, & BergerJonathan. (1995). A Model of Performance, Interaction, and Improvisation. Proc. Int. Computer Music Conference. International Computer Music Association.

HuronDavid. (2006). Sweet Anticipation: Music and the psychology of expectation. Cambridge, Massachusetts: MIT Press.

KeilCharles. (1966). Motion and feeling through music. The Journal of Aesthetics and Art Criticism, 24 (3), 337-349.
KeilCharles. (1966). Motion and feeling through music. The Journal of Aesthetics and Art Criticism, 24 (3), 337-349.
KeilCharles. (1987). Participatory Discrepancies and the Power of Music. Cultural Anthropology, 2 (3), 275-283.
KirkeAlexis, & MirandaR.Deuardo. (2013). An overview of computer ysstems for expressive music performances. Springer.
LemanMark, & NavedaLuiz. (2010). Basic Gestures as Spatiotemporal Reference Frames for Repetitive Dance/Music Patterns in Samba and Charleston. Music Perception: An Interdisciplinary Journal, 28 (1), 71-91.
Levitin, D. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. Perception & Psychophysics *, 927-935.*
LewisLowellJ. (1995). Genre and embodiment: from Brazilian capoeira to the ethnology of human movement. Cultural Anthropology, 10 (2), 221-243.
LimbJ.Charles, & BraunR.Allen. (2008). Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation. PLoS ONE, 3 (2), e1679.
MasonH.Paul. (2013). Intracultural and Intercultural Dynamics of Capoeira Training in Brazil. Global Ethnographic, 1 (1), 1-8.
McPhersonMalinda, & LimbJ.Charles. (2013). Difficulties in the neuroscience of creativity: jazz improvisation and the scientific method. Ann. N.Y. Acad. Sci.(1303), 80-83.
Repp, B. (1999). A microcosm of musical expression. i. quantitave analysis of pianists' timing in the initial measures of chopin's etude in e major. Journal of Acoustical Society of America *, 1972-1988.*
Repp, B. (1995). Expressive timing in schumann's traumerei : An analysis of performances. Journal of Acoustical Society of America *, 2413-2427.*
Repp, B. (1995a). Expressive timing in schumann's traumerei : An analysis of performances. Journal of Acoustical Society of America *, 2413-2427.*
Repp, B. (1995). Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: Some Percepeptual Evidence. Music Perception: An Interdisciplinary Journal *, 39-57.*
Sprio, N., Gold, N., & Rink, J. (2010). The form of performance: analyzing pattern distribution in select recordings of chopin's mazurka op. 24 no. 2. Musicae Scientiae *, 23-55.*
Talmon-ChvaicerMaya. (2008). The Hidden History of Capoeira: A collision of cultures in the Brazilian battle dance. Austin: University of Texas Press.
TaylorGerard. (2005). Capoeira: The Jogo de Angola from Luanda to Cyberspace (卷 1). Berkeley, CA: North Atlantic Books.
Widmer, G., & Tobudic, A. (2003). Playing Mozart by Analogy: Learning Multi-level Timing and Dynamics Strategies. Journnal of New Music Research *, 259-268.*
Widmer, G., Flossmann, S., & Grachten, M. (2010). YQX plays Chopin. AI Magazine *, 23-34.*
WillsonMargaret. (2001). Designs of Deception: Concepts of Consciousness, Spirituality and Survival in Capoeira Angola in Salvador, Brazil. Anthropology of Consciousness, 12 (1), 19-36.

[3-20] IDENTIFYING STRUCTURE IN A (16,9) HYPERDIATONIC SYSTEM

Christopher Bartlette[1], Anson Jablinski[2]
[1]Binghamton University, USA, [2]University of Houston, USA
Poster

We investigated whether listeners' judgments of tension in a novel hyperdiatonic system—with 16 chromatic tones equally dividing the octave, 9 tones in the diatonic scale, and tetrads that are maximally even sets in the diatonic scale—would produce results similar to tension ratings for the usual diatonic scale in Bigand, Parncutt, and Lerdahl (1996). Twenty-four participants (non-AP) listened to 64 trials, with each trial containing three chords: a tonic tetrad, a test tetrad, and the tonic tetrad repeated. Each trial was preceded by a 6-tone melody drawn from the diatonic scale and centered around the tonic tone. Each chord—composed of four sine tones—was 1.5 s long, separated by 0.5 s of silence. The test tetrads included all of the possible transpositions of the four types of diatonic tetrads; participants rated the tension produced by each test tetrad on a 7-point scale. The results were similar to those in Bigand et al.: Diatonic chords were less tense than non-diatonic chords; chords analogous to tonic, subdominant, and dominant were less tense than the other diatonic chords; and the tonic chord was less tense than the subdominant and dominant chords. Also, there was a correlation between ratings and tonal pitch space distances (based on Lerdahl, 2001) for diatonic chords and for the two most common diatonic tetrad types. Thirty-seven participants (non-AP) took part in a follow-up study, with the same method as the first study; however, the chords were presented in a lower register, and the distance the bass line moved from tonic tetrad to test tetrad was reduced (as this had an effect on ratings in the first study). The results were the same as in the first study; in addition, there was a correlation between ratings and tonal pitch space distances across all of the test tetrads. We identify qualities of hyperdiatonic structures that may have influenced the participants' responses, and we suggest the possibility of a generalized cognitive preference for diatonic and hyperdiatonic structures.

[3-21] Analysis on the method and effectiveness of learning piano of a child with autism

Shuo Chen[1]
[1]Shuochen, China
Poster

ABSTRACT

Based on the three theories of music therapy, applied behavior analysis and structured educational system, the author trained by playing piano a child with autistic individual for a period of four months. The author explores the method of learning piano of child with autism by doing this. What's more, explore the effectiveness of learning piano to improve the autistic 's mood and concentration and the interpersonal problems.

The paper is mainly studying piano teaching which using variety of methods:

1. Using the decomposition technique of applied behavior analysis to train autism's piano playing.

2. Using the visual cues of structure of education to learn the musical notation stave.
3. Using video method to consolidate student's learning and practicing piano effect and the method is also applied in practice concerto.
4. Promoting student's exercises by using family of integrated education.
5. Using a variety of forms to train play piano such as concerto with cello and violin and sing.

The author can draw a conclusion by the training and counseling video of the autism child:

1. The autism can improve the concentration by training piano.
2. The autism can improve the ability of cooperation by concerto and sing.
3. The autism can relieve the tension and enhance their self-confidence by learning piano.

Summary: the method of learning piano to autism child is different from the normal child. By this study, the author concluded a variety of methods to learn piano for children with autism and summed up the positive impact on child with autism who learns piano. The author hope that this research can make contribute to the scholars.

1. A BASIC CONDITION AND TREATMENT

Autism (also called Autism syndrome, Khans syndrome) is a disorder of neural development which involving perception, emotion, language, thought, action and behavior. It is explained by a description of medical diagnosis is that autism is characterized. By impaired social interaction, verbal and non-verbal communication, and by restricted and repetitive behavior (MingXie,2007). Beginning of the 21st century, the prevalence of autism is 13 million and is about per 100 children worldwide in china. Global autism has reached 67 million. In 2012, The China Disabled Persons' Federation show that there are 1.1 million people learned in the various agencies. With the number of people diagnosed with autism has been increasing dramatically, their education should be aroused great concern and attention. this paper selects an autism child which was diagnosed by the Sixth Hospital of Peking. Author based on his specific situation to develop treatment goals and treatment methods, the results were as follows:

2. 1INTRODUCTIONS THEBASICSITUATION

An child, male, March 4, 2003, 11years old. Autism(Asperser syndrome):

(1) His memory is beyond the ordinary especially for the melody. He is good at math. When he was a teens he can calculate the double-digit multiplication and can also say out the day and week of lunar calendar.

(2) He is serious emotional problems, but there is no aggressive behavior to others. He was learned in ordinary primary schools until he had been found out that it is impossible to integrate into normal peers. In September 2013, he is switching to a special education school and his mother study with him at school and at home. His parents are senior teachers and have stringent requirements to him. What's more, they hope him to have a better future which resulted in the nervous and bad temperament.

(3) He can not pay more attention on study and needs the teachers to guide.

(4) His language expression ability is poor and can not communicate well with his classmates.

(5) His coordinated ability is poor and he is not good at skip rope. According to the observation, His balance ability is good and he can practice the Bicycle by himself.

(6) Music learning is mainly composed of vocal music and piano. Since 2012, he insisted on sing and piano lessons at the training institutions every week. In2013, he through the seven tests when he participated in the vocal music test. The piano has reached the secondary level. Once a week, he can finish 6-8 tracks at least. (The use of teaching materials mainly include: children's Hanon, small Thomson, the big Thomson, the children's songs played piano tutorial, etc).

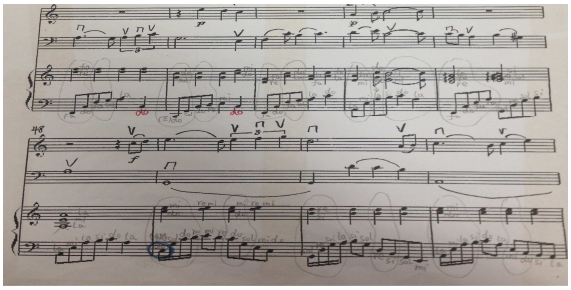
2.1. TREATMENT GOALSANDMETHODS

According to the basic situation of children, the author arranged the individual training for four month. Every week, he takes the piano class for 30 minutes. Professor main content is playing and singing the song of "invisible wings" and "sky city". In the process of study, the author mainly used the following methods:

(1) Used task decomposition of applied behavior Analysis technique in training of pianos studying: Applied behavioral Analysis (ABA) is a theory which was formed by the University of California, Los Angeles. It is based on the theory and practice of behaviorism psychology. What's more, it was designed specifically to target method for the treatment of children with autism. The core part of the treatment is task decomposition technique (Discrete 'Trial Therapy DTT) .By analysis and resolving the task, it break down the complex training into smaller units, solved them one by one. The decomposition method of treatment goals is not only reduced the difficulty of learning but also promote the learning motivation. Make the autism to experience success and to actively cooperate with training and learning behavior at the same time (Yan Zhang,2006).

Applying this method to the children's piano learning is mainly because he could not recognize the staff. And it is a huge task for him to master a large number of staff. So the author used the method that Playing a phrase and let him to imitate, and then through the two phrases series study, until he can finish learning a passages or even more. By Imitation playing and repeated practice, he complete the whole song.

(2) Applying structured education of visual cues to the learning of staff. In the daily learning of autism, a large number of visual cues are helpful for their learning. For example, the teacher will post out colorful curriculum and the new words in the classroom to help them understand and memory. The child is difficult in the studying of staff. In former piano learning, he can only sing "do re mi far sol la si". The author used the method which noted the roll-call annotation to help him to practice at home. And marking some changes with different color to help him better memory melody.



Picture 1: This is a clip of concerto "Sky City" which can be seen that the author used a lot of visual cues in the process of learning, such as: it is marked roll in the each played tone. The red mark where it is easy to play the wrong tone, the blue mark where it is prompted him not to play an octave lower and the right-hand man with a pencil on the circle indicates the location and position of the sound range.

(3) Using the method of video to consolidate the effect of student learning and practice the piano, and apply this method to the concerto practice: in order to better guide the children to learn piano, the author also points on each piece passages play the demonstration of break up, hands segmentation play a demonstration and the overall demonstration, and recorded it, for he exercises every day. And, with practice the violin concerto, also at the scene of the cooperation time of them live recorded, so that children are familiar with and to practice in class or home.

(4)Improving the studying of students by using family inclusive education: his mother accompany with his study every time. So his mother has a good knowledge about the training acquirement and teaching progress. It is helpful for him to practice at home.

(5) Adopting many kinds of form to study piano: the practice is not only used in solo play. It is mainly taking considering that the child is not good at communication and he likes to play the piano and sing with his song. So the author designs two ways to teach. The one is to teach him play with his song and the other is to play with violin.

3. TREATMENT RESULTS

The autism child can finish the concerto "Sky city" after nearly four months training. In December 3,2013 ,he was successful in the large-scale theater perfomances in Beijing. And he also completed the piano to sing "Invisible Wings". In December 31, 2013, he participate in the school celebration New Year's activities. By analyzing the weekdays a training video and video performances, It can be obtained the following results:

(1) Training piano to improve his attention:

First, when a pre-training test, the children can complete the "Little Red Riding Hood," "Red River Valley" and "Moscow Nights" and other songs playing short (about 40 seconds per song); And after four months of training students can play the concerto back "Sky city" (approximately 4 minutes and 30 seconds) and "Invisible Wings" (about 2 minutes and 20 seconds), and in the course of play can maintain full concentration.The author found that the autism had a better attention for he did not be affected by his mother's behavior during he is playing "Sky City" which left his seat after the 13 bar and appeared at the autism's behind at the end of 71 bar .What's more, When he was playing 38 bar, there was a little girl to dance through the play who was watching him to play the left 3 bar. A conclusion can be draw that this training increases the attention span of children.

Secondly, in the course of finishing the training, the autism never follow the therapist directive and needed to put a requires each time. Gradually, he developed into a mold that can be completed the task in accordance with the instruction immediately. And he gradually developed into paying more attention to the therapist behavior. It can be proved that this training improves his ability to transformed from unintentionally into conscious attention.

(2) The autism strengthen the ability to communicate with others by performed the concerto and playing and singing it by himself.

First of all, the finishing of "Sky city" is inseparable with the cooperation with others. For the autism , he usually plays the songs with strong melodic . So it is a huge challenge for him to play the concerto which is mainly consist of broking chords of the accompaniment. As the author can see from the video, at the beginning, the autism had a bad temperament for he can not find the melodic and it is prominent of his bad moon. However, after the constant communication with him, he gradually adapted to the boring music without melodic tunes. By this process, the author find that the autism can be established the effective communication with his parents and therapist.

Secondly, the concerto "Sky City" is playing in the form of the autism playing the piano part and two middle school students played the cello and violin, they played together. Before rehearsals, it is just arranged an ensemble practice. When the first pass ensemble practice, there have been cases of the autism playing in the chaos and bombs and he can not find the music score to the middle part of the course. But in the second time ensemble, he will be able to adapt to this form under the guidance of the author. After that, he can make use of the recorded video concerto to practice and achieve a good understanding. To the rehearsal time, when he played the last page (65-73 bars) on the chords, he also tied to observe the progress of two students and knew how to wait for them to put a measure before the completion of the melody played simultaneously pull into the next section of play. This concerto training used in the subtle language of music taught children how to approach this work with others and it also can develop the ability to cooperate with others.

Third, by singing and playing "Invisible Wings" , it contributed to the autism to use music to express and to increase interaction with teachers and students in his school. What's more, it also can meet the desires of his performances.

(3)The piano training can relieve his tension and enhance his self-confidence: The autism acquired a sense of satisfaction seeing from the process of his moon changing from angry to enjoyment and also can be seen by he finished Concerto "Sky city" and sing "Invisible Wings" got applause. Through music learning, it eased some of his negative emotions and it enhance his self-confidence by performance.

Table 1: As we can see from the table, the autism can cooperate with others generally well after three times training. After five times practice with live video, he can gradually finish the performance well for his change from tension to relax and repeatedly to smile. After a lot of video practice, he had no fright to rehearsal and performance and he finished the show very well.

The practice forms of playing the concerto	Practice times	The effect
Practice on site	1	3 times mistake
	2	He can cooperate with others generally well under the teacher's guidance but he has some emotional problems.
	3	He can cooperate with others generally well under the guidance but there are some loud noisy in the process of playing.
Practice by watching the video	1	He can't keep up with the whole paragraphs under the condition of noninterference.
	2	There is a situation which he threw the spectrum section but can keep up with it in a timely in the case of intervention.
	3	He can ensemble with others without intervention.
	4	He can ensemble with others without intervention and he can express his satisfied.
	5	He can ensemble with others without intervention and show a smile with relaxed.
Rehearsal	1	He can ensemble with others very well without intervention.
	2	He can ensemble with others very well without intervention.
Performance	1	He can ensemble with others very well without intervention.

4. DISCUSSION

By piano training, the author summarizes some piano learning method for children with autism. such as: making use of a large number of visual cues can help him learn to read music and memory. Making a large of demonstration in the skills learning for they have a good ability to mimic and it should be used to break one by one and then the song exploded integrated approach. It is good to practice integrity and concerto when the demonstration video recorded down to let him practice or play the concerto with the effective exercise. In case of a conditional, it is necessary to allow parents to participate in training. It not only can reduce the tension and discomfort of children, but also played a role in auxiliary exercises after class. Using a variety of exercises can enhance children's interest and also to strengthen the ability to cooperate and communicate with others.

Of course, after the piano training, I also found some shortcomings:

(1) There are some limitations in terms of the parents involved in the training. The author can not control the strong wishes of the parents and the interference to students. In a degree, it not only affect the training effect but also affect the communication between the author and the children.

(2) During to some activity in the school, the therapist can not guarantee a daily training schedule every week, which has bad effect to the autistic children emotional stability and adaptability.

(3) In the process of studying stave, the author intent to exercise recognition the stave interspersed with exercises and do not want to rely recognize the roll-call approach. But after the first implementation, the author found that it can cause the bad feelings of the child. So the author did not insist on this kind of practice. The author hopes to try to teach children to learn by using a small card about the stave and making full use of the methods of visual cues.

(4) In the form of piano learning can also be expanded. For example, playing in four hands and to the accompaniment for others and so on.

In short, by training autism child to play piano, the author find out that it has some positive impacts on improving concentration and relieving emotional aspects of communication and cooperation with others. Of course, this study is only applied to a case autism. So some of the inductive method can be generalized but it should be adjusted according to different characteristics of children with autism. But the author learned a lot through this research and hopes to in do more in-depth and detailed study in the future. The author also hopes that this study can provide a valuable reference point for autistic children to learn piano.

5. REFERENCES

谢明.《孤独症儿童的教育康复》.天津教育出版社.2007. 3-7.

世界自闭症日.http://baike.baidu.com.

自闭症儿童数量快速增长，已占中国精神残疾首位.中国青年报.2013. http://money.163.com/13/0403/09/8RHCOEQU00253B0H.html.

张焱.孤独症儿童音乐治疗过程中若干问题的探讨.中央音乐学院学报.2006 年第 2 期.

[3-22] USING DISCRETE CALCULUS TO CORRELATE A SINUSOIDAL MODEL OF METER WITH NORMATIVE RHYTHMS AND STRUCTURAL ACCENTS

Scott Murphy[1]
[1]*University of Kansas, USA*

Poster

As a prologue to an empirical study of syncopation in American popular music, Huron and Ommen (2006) discuss the difference between the conventionally accepted sixteen-fold distribution of metrical weight in a 4/4 measure, which is symmetrical around the first and third beats (for example, the second and fourth beats have the same metrical weight) and the frequency of note onsets at each of these sixteen positions in a representative corpus of monophonic music, frequencies which are asymmetrical around the first and third beats (for example, more onsets occur

on the fourth beat than the second). In hypothesizing how listeners attend metrically, the authors evince a preference for the use of a set of learned rhythmic schemas, which relates to the second structure, rather than the coordination of multiple metric oscillators, which relates to the first. While the authors recognize similarities and differences between these two structures, they do not explore these similarities and differences further. The present study proposes a precise relationship between these two structures that takes advantage of the multiple-oscillator approach. Large and Palmer (2002) model metrical hierarchy through the combination of multiple sinusoidal functions of equal amplitudes but different frequencies into a complex metrical function. A combination of five such waves, each twice the frequency of the next slower, well models the first structure described above. However, the discrete derivative for each of the sixteen equally distributed points well matches with near-perfect correlation the second structure described above. In other words, the degree of change of metrical weight from one moment to the next directly corresponds to the likelihood of an onset, if metrical weight of a moment derives not from a discrete hierarchy but instead from a continuous function of the sum of the underlying metrical periodicities as simple sine waves. Expanding the purview from a single measure to a phrase, an equivalent continuous and composite metric waveform that stretches over four measures, when combined with a lateness preference, strongly predicts the location for normative points of cadential arrival.

[3-23] HAND REHABILITATION USING MIDI KEYBOARD PLAYING IN ADOLESCENTS WITH BRAIN DAMAGE: A PRELIMINARY STUDY.

Hyun Ju Chong[1], Sung-Rae Cho[2], Soo Ji Kim[1]
[1]Adepartment of Music Therapy, Graudate School and Ewha Music Rehabilitation Center, Ewha Womans University, Seoul, Korea, Korea, Republic of, [2]Bdepartment and Research Institute of Rehabilitation Medicine, Yonsei University College of Medicine, Seoul, Korea, Korea, Republic of
Poster

1. Background: Recent research regarding implementing music instruments for physical rehabilitation is increased; yet, adolescents with brain damage are less focused. As a sequential, programmed movement of fingers, keyboard playing is promising to implement for inducing execution and a high level of coordination during finger movements. Keyboard playing, as an option for physically and emotionally intriguing for adolescents in rehab setting, is in need to investigate.

2. Objective: The purpose of this study is to evaluate the effects of keyboard playing using Musical Instrument Digital Interface (MIDI) on finger movement exercise for adolescents with brain damage.

3. Methods: Eight adolescents with brain damage, aged 9 to 18 years (M = 13 years, SD = 2.78), in physical rehabilitation settings participated in this study. Measurements included MIDI keyboard playing for depressing force of the fingers and hand function tests (Grip and pinch strength, Box and block test (BBT), and the Jebsen Taylor hand function test).

4. Results: Results showed the increased velocity of all fingers on the MIDI-based test, and statistical significance was found in the velocity of F2 (index finger), F3 (middle finger), and F5 (little finger) between pre and post-training tests. Correlation between the depressing force of the finger and hand function tests was also analyzed. A strong positive correlation between the measure of grip power and the depressing force of F2 and F5 on the Grip and pinch strength, also all fingers showed strong correlation between MIDI results and BBT. For the Jebsen Taylor Hand Function test, only the moving light objects task at post-training yielded strong correlation with MIDI results of all fingers.

5. Conclusions: The results imply the possibility of using keyboard playing for hand rehabilitation, especially in the depressing force of individual finger sequential movements. Further investigation is needed to define the feasibility of the MIDI program for valid hand rehabilitation for people with brain damage.

[3-24] THE EFFECT OF ENSEMBLE SIZE, MUSICAL ROLE, AND INDIVIDUAL PERFORMER ON COLLECTIVE ENTRAINMENT IN KHASONKA DRUMMING FROM MALI

Justin London[1], Nori Jacoby[2], Rainer Polak[3]
[1]Carleton College, USA, [2]Hochschule für Musik Und Tanz, Köln, Germany, [3]The Hebrew University of Jerusalem, Israel
Poster

1. Background: Maintaining synchrony in a musical ensemble requires each musician to constantly adjust the timing of their actions to those of their fellow musicians. These responses can be understood as a form of mutually adaptive error correction (Repp 2005, Repp & Su 2013, Himberg 2014). Iterated cross-correlation of timing asynchronies may be used as measure of coordination and mutual entrainment amongst ensemble members (Keller 2007; Moore & Chen 2010; Wing, et al. 2011; Jacoby & Repp 2012).

2. Aims and Method: To show the influences of ensemble size, musical role, and specific performers on ensemble synchrony, data were collected from four performances of Khasonka drumming in natural settings. These instruments (drums and bells) produce crisp onsets yielding clean timing data. Each drum or bell has a clear role: a basic timekeeper, a characteristic melorhythmic figure, and an improvisational lead drum. Our data also involved performances (a) with and without an extra drum, and (b) where two players swapped lead and accompaniment parts. From the collected data we used multivariate regression to predict the subsequent onsets of each instrument from the eight onsets preceding it in the other parts. The percentage of explained variance of this prediction, compared with the total variance of the target, serves as an index of the degree of entrainment of the target to itself and to each other part, expressed as a set of paired comparisons.

3. Results: As expected, the timekeeper bell plays a dominant role in all contexts, but is not always the strongest influence on every other instrument. The addition of an extra drum changes the couplings in the other parts. Performer-specific differences in the dominance of and sensitivity to other ensemble members' timings, made evident when performers switched musical roles, were found.

4. Conclusions: Malian drumming involves non-isochronous rhythms and significant tempo changes (Polak 2010, London 2012, Polak & London 2014), yet exhibits patterns of error correction similar to those in western classical music ensembles (Wing et al 2013). The interaction between individual sensitivity and culture/genre-specific performance role may be most evident in a natural performance setting.

[3-25] COMMUNICATING NATIONAL AND TEMPORAL ORIGIN OF MUSIC – AN EXPERIMENTAL APPROACH TO APPLIED MUSICAL SEMANTICS

Hauke Egermann[1], Cornelius Stiegler[2]
[1]Audio Communication Group, Technische Universität Berlin, Germany, [2]Nhb Corporate Sound, Germany
Poster

1. Background: In marketing and media music is often selected in order to communicate extra-musical meanings. However, it remains unclear, how specific and reliable, those extra-musical associations can be elicited in listeners, potentially causing unsuccessful decoding of the intended meanings.

2. Aims: Based on two audio branding use cases, we therefore tested in two web experiments, whether music, selected to communicate information about its national and temporal origin, was able to evoke the intended associations. Experiment 1 tested for correct recognition of musical decades with music excerpts each chosen to represent one out of 12 decades from 1900 to 2010, and Experiment 2 tested for correct recognition of 12 countries associated with music recordings.

3. Methods: In Experiment 1 89 participants (age M=30 years) listened to 12 musical excerpts in random order. Subsequently, they indicated which decade they associated with the musical stimuli. They were also asked to list 5 free associations with the stimuli presented. In Experiment 2 140 participants (age M=35 years) listened to 10 instrumental music excerpts chosen to represent 10 different regional music styles from 10 European countries. Here, 55 participants were presented a map of Europe and asked to click on the country they thought that excerpt originated from; avoiding priming effects, all other participants were asked to enter up to 5 different association terms in an open text field.

4. Results: In Experiment 1, results indicated that for most excerpts (except between 1920-1940), decades were correctly associated. Experiment 2 indicated that the capacity to associate the correct country with a nation varies extremely: while more than 90 % of participants correctly recognized the excerpt from Spain, recognition rate for Italy and Portugal was below chance level.

5. Conclusions: Taken together, these results indicate, that music may activate shared meaning structures that could be used for communication purposes. However, the success of these measures depends on the similarity of inter-individual extra-musical association networks and the strength of learning of associations between music and other features (exemplified by the low recognition rate of some countries). The results will be discussed with respect to the representativeness of sampled stimuli and participants.

Author Index

A

Abrams, Daniel A. 177
Adachi, Mayumi.....35, 213, 364, 365
Adzei, Senyo 208
Agres, Kat..... 239
Ala-Ruona, Esa 214
Albrecht, Joshua..... 128
Alcalde, Bruno 80
Allen, Erin 148
Alluri, Vinoo 396
Altenmueller, Eckart..... 177
Ambrazevičius, Rytis..... 255
Anzenbacher, Christoph 356
Arnau, Eric 320
Arthur, Claire 194
Arzt, Andreas 177
Ashley, Richard..... 105, 204
Avantaggiato, Massimo 41
Avron, Amit..... 146

B

Bååth, Rasmus.....71
Bae, Byung-Chull 263
Bae, Dahye 397
Bailes, Freya.....41
Ballús, Andreu..... 320
Bao, Zhimin..... 146
Barlow, Christopher..... 256
Bartlette, Christopher..... 411
Basiński, Krzysztof..... 377
Bauer, Anna-Katharina Raphaela 69
Berger, Jonathan... 52, 178, 214, 245, 304
Bisesi, Erica.....76, 124, 248
Bishop, Laura 64
Black, Dawn 250, 405
Black, Justin 342
Bodinger, Marlies 76
Bogert, Brigitte..... 396
Bountouridis, Dimitrios 238
Bouwer, Fleur..... 284
Brattico, Elvira 70, 184, 396
Broughton, Mary..... 184
Brown, Jenine L. 133
Brown, Steven 205
Büdenbender, Niklas..... 157
Burger, Birgitta 12, 214, 395
Burgoyne, John Ashley..... 238
Burunat, **I**balla..... 70, 396

C

Caplin, William E. 239
Cardoso De Araujo, Rosane 368
Cardoso, Edgar.....32
Carlyon, Robert..... 336
Caron, Sylvain..... 124
Cazabon, Dominique Beauregard 276
Chandra, Rohan..... 177
Chang, Son-Hwa 284
Chang, Sungkyun..... 271
Chen, Ao..... 171
Chen, Chung-Ping.....65
Chen, Jyh-Horng 154
Chen, Lingjing..... 365
Chen, Shing-Jen.....364
Chen, Shuo.....411
Cheong, Yong Jeon 148
Cheong, Yun-Gyung 263
Chew, Elaine.....405
Cho, Sung-Rae.....415
Chon, Song Hui 185, 289
Chong, Hyun Ju 127, 415
Chuen, Lorraine 106
Chung, Chun Kee 157, 397
Chung, Heewon 133
Chung, Hye-Yoon.....145
Clarke, Eric 158, 172
Cohen-Shani, Neta.....396
Collins, Tom 177
Condit-Schultz, Nathaniel.....42
Coorevits, Esther 12
Copeland, Evan.....259
Cross, Ian.....342
Currie, Thomas 205
Cusack, Rhodri 336
Czedik-Eysenberg, Isabella.....21, 356

D

Dalla Bella, Simone.....176
Davidson-Kelly, Kirsteen.....182
Davies, Kathryn 4
Dawber, Emily J. 205
De Andrade, Margaret Amaral 201
De Araujo, Rosane Cardoso.....201
De Francisco, Martha 20
Dean, Roger 41
Demorest, Steven.....213
Deutsch, Diana.....190
Devaney, Johanna.....126
Di Bernardo, Cristina 337
Di Stefano, Nicola 32
Dmochowski, Jacek.....178
Douglas, Chelsea 276, 342

E

Eckl, Florian.....248
Eerola, Tuomas.....259
Egermann, Hauke 11, 183, 416
Eitan, Zohar.....396
Elvers, Paul 18
Ericsson, Margareta.....222

F

Fabian, Dorottya 183
Farbood, Morwaread 22
Farley, Alison.....70
Farrell, Helen 228
Farrugia, Nicolas 176, 336
Fine, Philip 113, 256, 333
Fischinger, Timo.....18, 19
Floridou, Georgina 148
Floridoy, Georgia.....336
Flowers, Paul 112
Foerstel, Alexander.....183
Font, Frederic.....320
Franek, Marek.....109
Fregoneze, Carmen.....398
Friberg, Anders.....248
Fuhrmann, Wolfgang 18
Fujihira, Haruna 69
Fujinaga, Ichiro 263
Fujioka, Takako..... 209
Fujiyama, Saki..... 233
Fulford, Robert 158

G

Garniez, Nancy..... 405
Gillard, Jessica..... 21
Ginsborg, Jane 158, 256
Gjerdingen, Robert..... 146
Godøy, Rolf Inge..... 46
Goebl, Werner 64, 164
Goldemberg, Ricardo..... 113
Goldman, Andrew 41
Golka, Gerald 164
Goodchild, Meghan..... 276
Graves, Jackson..... 377
Groarke, Jenny.....113, 352
Gruhn, Wilfried 86

H

Hadley, Lauren 173, 245
Hamamura, Mariko 297
Hameem, Hamzeen 214
Han, Yo-Jung..... 353
Handa, Yasushi..... 365
Hankinson, Andrew 263
Hansen, Kjetil F..... 385
Hansen, Niels Chr 147
Harlow, Randall 164
Hartmann, Martin..... 165
Harvey, Dominic 184
Hasuo, Emi 69
Hattori, Yuko..... 9
Haugen, Mari Romarheim 46
Hawkins, Sarah..... 342
Henthorn, Trevor..... 190
Herrmann, Christoph..... 69
Hill, Niklas..... 21
Himberg, Tommi.....90, 176, 344
Hiraga, Rumi 348, 385
Hirano, Takeshi 405
Hogan, Michael 113, 352
Hohagen, Jesper 352
Hong, S. 182
Hong, Sujin..... 69
Honing, Henkjan 172, 284
Hopkins, Carl..... 158
Hoshi, Nobuko..... 364
Hoshino, Yumi 228
Huang, Yun-An 154
Huberth, Madeline..... 342
Huo, Xinyang 364
Huotilainen, Minna 19
Huron, David 107, 185, 259, 289, 301
Hwang, Sung Hyun 365

I

Ichikawa, Makoto..... 316
Imamura, Hidetaka..... 311
Ingvarsdottir, Kristin Osk 71
Irie, Toshihiro..... 395

Ishikawa, Daisuke 239
Iwamiya, Shin-Ichiro 96, 233, 297

J

Jablinski, Anson 411
Jacoby, Nori..... 415
Jakubowski, Kelly..... 171, 336
Janata, Petr 373
Järvinen-Lepistö, Päivi 214
Jiang, Cong..... 190
Jung, Bernhard 222

K

Kager, Rene 171
Kahlhoefer, Daniel.....11
Kajikawa, Sachiyo 357
Kamekawa, Toru..... 311
Kaneshiro, Blair 52, 178, 214, 245
Kaneuchi, Masanori 365
Karavellas, Themistoklis 238
Katou, Hisataka 398
Kawakami, Hiroshi 369, 395
Kaya, Mustafa Uğur..... 404
Keller, Peter.....9, 56, 175, 245
Kelley, Jamey 213
Kiilu, Kristi.....86
Kim, Ji Chul 158, 164, 185
Kim, Jin Hyun 109
Kim, June Sic 157, 397
Kim, Jung Nyo 158
Kim, Ki-Hong 96
Kim, Soo Ji.....127, 147, 415
Kim, Sungyoung 311
Kim, Youn..... 110
King, Richard 311
Kinoshita, Hiroshi 405
Kirby, Lucy 333
Kishida, Takuya 20, 69
Klyn, Niall 148
Koenig, Rasmus 9
Koizumi, Nobuo..... 239
Kojima, Yuriko Hase 277
Kopiez, Reinhard 201, 205, 232
Krause, Amanda.....10
Kreutz, Gunter.....69, 112, 157
Kringelbach, Morten..... 208
Kueter, Kristin..... 245
Kusanagi, Emiko..... 364
Kuubeterzie, John 208
Kuwabara, Hiroshi 382
KWAK, Jun Young..... 214

L

Lami, Dominique 396
Lappe, Claudia 396
Lappe, Markus..... 396
Large, Edward.....158, 164, 185
Larsson, Karina 222
Lartillot, Olivier 80, 165
Lawendowski, Rafał 377
Lee, Eun-Jeong..... 386
Lee, Ji Yeon.....65
Lee, Kyogu 271
Lee, Kyung Myun 12, 157, 397
Lee, So-Young 196
Lee, Sukjae 4
Lee, Sun Hee.....332
Lehmann, Andreas C.....201
Leman, Marc.....128, 248
Lembke, Sven-Amin 20
Lenoir, Matthieu 128
Lerud, Karl 185
Lesaffre, Micheline.....128
Levine, Scott 20
Li, Ma 250
Li, Ming.....373
Li, Shengchen 405
Li, Xiaonuo 171
Liebscher, Frank 357
Lien, Yunn-Wen 154
Limb, Charles.....343
Lindborg, Permagus 26
Liu, Jingyi 208
Liu, Xiaoluan 397
London, Justin..... 176, 386, 415
Lopez-Gonzalez, Monica 343
Lothwesen, Kai Stefan 104
Lu, Xuejing 188
Luck, Geoff 12, 395
Lui, Simon.....214

M

Macdonald, Raymond 112
Madison, Guy.....65
Makam, Megha 52, 214
Manning, Fiona.....75
Margulis, Elizabeth..... 79, 106, 342
Martinez, Diana 106
Marui, Atsushi 267, 311
Marumo, Mieko 395
Matsubara, Masaki 385
Matsuda, Shu 348
Matsumoto, Junko 343
Matsuzawa, Tetsuro.....9
Mcadams, Stephen26, 153, 183, 239, 276, 342, 373
Mcadams, Stephen.....20
Mccullough, Stephanie.....332
Mcpherson, Malinda.....343
Meals, Cory.....394
Menon, Vinod 177
Miell, Dorothy 112
Mills, Peta 56
Minakami, Hisanori.....365
Misenhelter, Dale.....320
Mito, Yuki.....369, 395
Miura, Masanobu117, 326, 369, 382, 387, 398
Miyamoto, Yuri 386
Miyawaki, Satoshi 398
Moelants, Dirk 12, 248
Moran, Nikki.....245
Morcom, Alexa 69, 176
Morgan, Jill 112
Mori, Fumiya 96
Morimoto, Satoshi 69, 170
Moriuchi, Hideo.....357
Morrison, Steven 394
Mosher, David 110
Mühlhans, Jörg 357
Müllensiefen, Daniel ...75, 146, 171, 271, 336
Mungan, Esra..... 404
Murphy, Scott..... 414
Musil, Jason 75

N

Nakajima, Yoshitaka..... 20, 69, 146, 170, 243
Nam, Hyun Sung..... 319
Nandi, Asoke 396
Nemoto, Iku 319
Neuhoff, Hans 19
Nieto, Oriol..... 320
Nijs, Luc..... 248
Nishimura, Akira..... 239
Nisula, Kirsten..... 259
Noble, Jason..... 276
Norcia, Anthony 178
North, Adrian..... 10

O

Obata, Satoshi..... 405
Odahara, Yuko..... 228
Oehler, Michael.....21, 356, 357
Ogawa, Yoko.....201, 337
Ogura, Sanae..... 213
Ohsawa, Chie..... 405
Oishi, Tadashi..... 364
Okada, Sota..... 382
Okazaki, Satoshi..... 316
Otsuka, Kazuhiko..... 348
Overy, K. 182
Overy, Katie.....69, 176
Oxenham, Andrew 377
Ozawa, Toru 395

P

Panebianco-Warrens, Clorinda..... 113
Pang, Ming Wai..... 277
Pantev, Christo 396
Park, Elizabeth..... 196
Park, Hye Ji..... 147
Park, Hye Young..... 104
Park, Hyekyung..... 353
Park, Jeong Mi..... 157
Parncutt, Richard.....76, 124, 248
Paulin, Johan..... 65
Pearce, Marcus 147, 239, 293
Peltola, Henna-Riikka 259
Pfordresher, Peter 213
Pickering, Martin..... 173
Pileri, Stefania 337
Platz, Friedrich 201, 205
Plumbley, Mark D..... 405
Podlipniak, Piotr..... 4
Podolak, Olivia M. 238
Pogrzeba, Loreen..... 222
Polak, Rainer 19, 386, 415
Price, Harry..... 394
Price, Khen 22
Puoliväli, Tuomas 70

Q

Quinn, Ian 320

R

Randall, William 19, 387

Rankin, Summer 343
Remijn, Gerard 311
Remijn, Gerard B. 20, 69, 170, 243
Reuter, Christoph 21, 356, 357
Reybrouck, Mark 100
Rezny, Lukas 109
Rickard, Nikki 19, 387
Ristaniemi, Tapani 70, 396
Ristmägi, Reet 86
Roberts, N. 182
Roberts, Neil..... 69, 176
Roncaglia-Denissen, M. Paula ... 172, 284
Rose, Jens-Peter 386
Rusch, René..... 153
Rush, Chloe 344
Russo, Frank..... 178

S

Saarikallio, Suvi 12, 19, 395
Sadakata, Makiko..... 147
Sakai, Emi..... 205
Sakurai, Mio 387
Sams, Mikko..... 70, 396
Sasaki, Takayuki 243
Sauve, Sarah 293
Savage, Patrick..... 205
Schaefer, R.S. 182
Schaefer, Rebecca S..... 176
Schellenberg, E. Glenn 205
Schellenberg, Glenn..... 356
Schmuckler, Mark A..... 238
Schubert, Emery..... 183
Schultz, Christof M. 11
Schutz, Michael.....21, 75, 106
Scirea, Marco 263
Sears, David..... 239, 263
Seedorf, Marten.....11
Seifert, Uwe..... 220
Seiffert, Gary..... 158
Seno, Takeharu..... 394
Seol, Jaeho..... 157
Shanahan, Daniel 106, 206
Shanahan, Eva 206
Shekar, Priyanka 209
Shinoda, Yukitaka..... 369, 395
Shiraishi, Kimio 69
Shoda, Haruka 35
Shoshi, Towaki..... 201
Siedenburg, Kai..... 26, 373
Skiada, Naoko 146
Slaney, Malcolm 304
Smedley, Eric 394
Song, Liming..... 373
South, Michael 271
Spence, Charles.....158
Speth, Florina.....220
Spiro, Neta 90, 344
Stevens, Catherine 105, 184
Stewart, Lauren.....148, 171, 293, 336
Stiegler, Cornelius 416
Stolzenburg, Frieder 27
Strange, Adam 267
Sturt, Patrick 173
Sun, Yanan.....188
Swaminathan, Swathi 356

T

Taga, Masae 365
Taher, Cecilia.....153
Täht, Karin 86
Takahashi, Junko 369
Takahashi, Keitaro.....348
Takamura, Hitoshi 364
Takishita, Fumiyuki.....233
Tekman, Hasan Gürkan.....90
Teng, Shan-Chuan 154
Terasawa, Hiroko.....385
Thompson, Marc..... 12, 158, 395
Thompson, Marc R. 176
Thompson, William Forde 188
Tian, Mi 250
Tillmann, Barbara 177
Timmers, Renee 183
Tobimatsu, Shozo 69
Toiviainen, Petri12, 70, 85, 165, 184, 395, 396
Tominaga, Hitoshi 386
Tomonaga, Masaki 9, 394
Tong, Xue.....361
Torrents, Alba G. 320
Traube, Caroline 124
Trehub, Sandra E. 205
Trochidis, Konstantinos 214
Tsai, Chen-Gia 65
Tsubaki, Yushiro 311
Tsuzaki, Minoru.....405

U

Uchida, Yukiko.....386
Ueda, Kazuo.....20, 69, 146, 243
Ueno, Satoshi.....319
Upham, Finn 127, 387

V

Van Balen, Jan 238
Van Beek, E. J. 182
Van Dyck, Edith 128
Van Noorden, Leon..... 109
Van Vugt, Floris T. 177
Vansteenkiste, Pieter..... 128
Veltkamp, Remco C..... 238
Von Georgi, Richard 104
Vuoskoski, Jonna 158, 259
Vuust, Peter 208

W

Wacker, Markus 222
Wahl, Michael 220
Wan, Morton..... 163
Wang, Li-Ching..... 11
Wang, Yuh-Wen..... 154
Webber, Jamie 276
Weidema, Joey L..... 172
Weis, Joachim 386
Weiss, Michael W. 205
White, Christopher 320
Wiering, Frans 238
Wild, Jonathan..... 276
Will, Udo 148
Williamson, Victoria.....148, 271, 336
Witek, Maria..... 208
Wolf, Anna 201, 205, 232
Won, Sook Young..... 304
Woszczyk, Wieslaw..... 311
Wu, Yi-Cheng 139

Y

Yahagi, Toru..... 233
Yamasaki, Mizuki 326
Yamasaki, Teruo 232, 386
Yamashita, Yuko..... 146
Yan, Yonghong 373
Yankera, Appiah Poku..... 208
Yazici, Z. Funda..... 404
Yi, Suk Won 157, 397
Yiu, Suki S. Y..... 277
Yoo, Ga Eul..... 127
Young, Christina 177
Yrtti, Antti..... 19
Yu, Lira..... 394
Yu, Szu-Pei..... 65

Z

Zhang, Pin..... 188
Zhang, Zhuo..... 90
Zhou, Shibin 190
Ziethen, Michael..... 21
Ziv, Naomi..... 10

ICMPC APSCOM 2014 SEOUL

TUESDAY CONCERT

2014. 08. 05. 20:30

Yun Ju-Yong Recital Hall, College of Music

Organized and moderated by **Sol LIM** (Seoul National University)

Lecture by **Dr. Sung Hwa Hong**

"Cochlear implant and music perception"

Video

"An introduction to The Snail of Love Association's Clarinet ensemble"

Performance

Clarinet solo | **Jung-Woo SON**

- You raise me up (with piano accompaniment)

Clarinet ensemble | **Jung-Woo SON, Chae-Jung HAN, Na-Yoon CHO**
Jae-Hyeok CHOE, Su-Jong CHOE (with bass clarinet and piano)

- I will / Let it be / Do Re Mi Song

INTERMISSION

Moderated by **You Jin KIM** (Yonsei University)

Performance

YI Ji-young

- Gayageum Sanjo (Seo Gong-chul School)



Dr. **Sung Hwa Hong**

1978.03-1984.02	M.D. College of Medicine, Seoul National University, Seoul, Korea
1987.05-1988.02	InternshipV
1988.03-1991.02	Residency, Department of Otolaryngology, Seoul National University Hospital, Seoul, Korea
1997.03-present	Professor, Department of Otolaryngology, Samsung Medical Center Joining Samsung Medical Center in 1993
2011.11-present	Executive Vice President, Research & Development Chief Research Officer, Samsung Medical Center
2013.05-present	President, The Korean Audiological Society



Professor **YI ji-young**

YI ji-young, one of the most representative young gayageum players in Korea, began her Korean music studies when she was five years old. She earned both the BA and MA in gayageum performance from Seoul National University and a Ph.D. at Ewha Womans University. She was the first Ph.D. in gayageum performance in Korean music history. She formerly served as a member the court music division of the National Center for Korean Traditional Performing Arts. Currently she is a Professor of Gayageum Performance in the Department of Korean Music in the College of Music at Seoul National University. She is a Candidate for Important Intangible Cultural Property No. 23, gayageum sanjo (solo instrumental folk genre) and gayageum byeongchang (folk song accompanied by gayageum). Since her professional debut, she has performed with numerous world-renowned orchestras and ensembles including Shanghai Orchestra, Tokyo City Philharmonic Orchestra, Jerusalem Philharmonic Orchestra, Atlas Ensemble, KNM Berlin Ensemble, etc. In addition, she has been the most frequently invited Korean musician at international music festivals including the Edinburgh Festival, MIDEM, ISCM, and Asian Composer's League. Her musical spectrum ranges from the very traditional to the extremely avant-garde. Colleagues and critics alike consider her to be the most talented gayageum artist of our time whose music represents a perfect harmony of the emotional and the sensible. She is also known as the most important figure in the interpretation and performance of contemporary gayageum repertoires.



Professor **Kim Woong-sik**

Kim Woong-sik is one of the most versatile percussionists.
He is an active member of CMRK and Puri and serves as an Adjunct Professor of National University of Arts and a lecturer of Seoul National University.



The Snail of Love Association, Inc.

The Snail of Love Association's Clarinet Ensemble was established in 2003, consisting of children and teenagers who had cochlear implantations due to their hearing impairment. The ensemble has built an extraordinary career with many activities: it has given concerts every year since 2005 and has been involved in several television shows, such as KBS Open Concerts in 2010 and KBS Family of Love in 2013. In addition, the ensemble has been invited more than 50 times to various venues such as National Assembly and medical conferences.

The Snail of Love
clarinet ensemble
since 2003



Clarinet **Jung - Woo, SON**

Jung-Woo Son, a 2nd year student at Seoul Dong-Buk High School, had cochlear implantation in 2004. He joined the Snail of Love Association's Clarinet Ensemble in 2008, with a 2nd degree of hearing disability.
At the National Students Music competition in 2011, Son was awarded with the special prize in string and woodwind part. More recently, he was nominated as the 3rd place in Clarinet part at the 2014 Seoul Orchestra competition.
Currently he is the leader of the Snail of Love Association's Clarinet Ensemble.



Sol Lim

Sol Lim is a doctoral student at Seoul National University majoring in music psychology.



You Jin Kim

You Jin Kim is a doctoral student in music and human learning at University of Texas at Austin. She holds her bachelor and master's degree in gayageum performance from Seoul National University and M. Ed. in music education from Yonsei University.

Service Information

REGISTRATION/HELP DESK

The registration/help desk is located in the lobby of Baekyang Hall, the main conference venue (see map on the back page). The desk will be staffed from 7:30AM to close for every day of the conference beginning on Monday August 4th.

BREAKFAST

We will provide free breakfasts from Monday August 4th to 8th except Wednesday in the lobby of Baekyang Hall. Breakfast will be served from 7:30AM to 8:30AM. We will offer doughnuts, fruits in season, coffee, and juice. We recommend taking advantage of free breakfasts.

LUNCH & DINNER

Participants are responsible for their own lunch and dinner arrangements. Please check out "Sinchon Map" for a list of restaurants – many within a fifteen minute walking distance from campus. A more detailed restaurant guide is available below.

ACCESS TO FREE WIRELESS INTERNET

Wireless internet access is available in the Baekyang hall. Just set up the wireless internet and install the access program by clicking 'Yonsei-info.' Once the program is successfully installed, double click the 'Yonsei wireless' icon. Input 'Yonsei_open' ID. You may now have access to the internet without password. Wireless internet can be accessed by your laptop computer, Iphone, Android phone, Macbook, and Blackberry.

PRINTING

Printing Service is provided by Yonsei University Library and Yonsei Co-op. Please check out below campus map. This service is available at the copy rooms on the B1F of Central Library and at the self copy corner on 2nd, 5th, and 6th floor of Yonsei-Samsung Library. You can use the printing service with networked printers on the B1F and the 6th floor Community Lounge of Central Library, Information Commons, 5th and 6th floor of Yonsei-Samsung Library.

- **Cost :** 60 won / page
- **Contacts :** Yonsei Co-op by (02) 2123-4039

Location		Services	Mon-Fri	Sat
Central Library	B1F Copy Room	Copy Room Hours Copy Room (Copy, print, bind)	09:00–17:30	
	2F Humanities Room	Self-copiers	09:00–20:00	09:00-17:00
	3F Social Sciences & Sciences Room	Self-copiers	09:00–20:00	09:00-17:00
	4F Humanites & Social Sciences Reference Room	Self-copiers, printers	09:00–20:00	09:00-17:00

COFFEE BREAKS

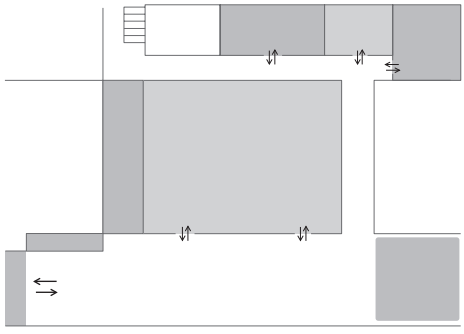
We will offer 4 coupons for free tea or coffee. You can use coupons at the Café booth in the lobby of Baekyang hall.

OPENING RECEPTION

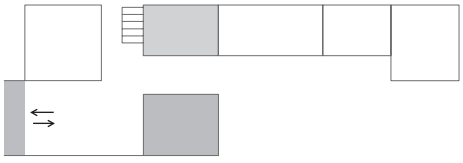
All delegates are invited to attend the opening reception on Monday August 4th from 12:00-1:00 PM on the Baekyang Hall's lobby. We hope to see all of you there.

Baekyang Hall Floor Guide

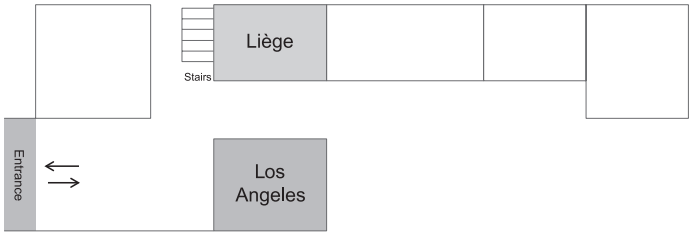
2F



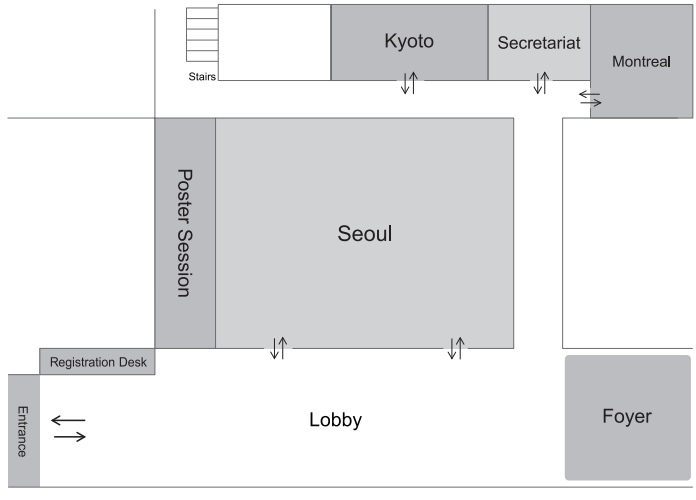
1F



Meeting Room 1F



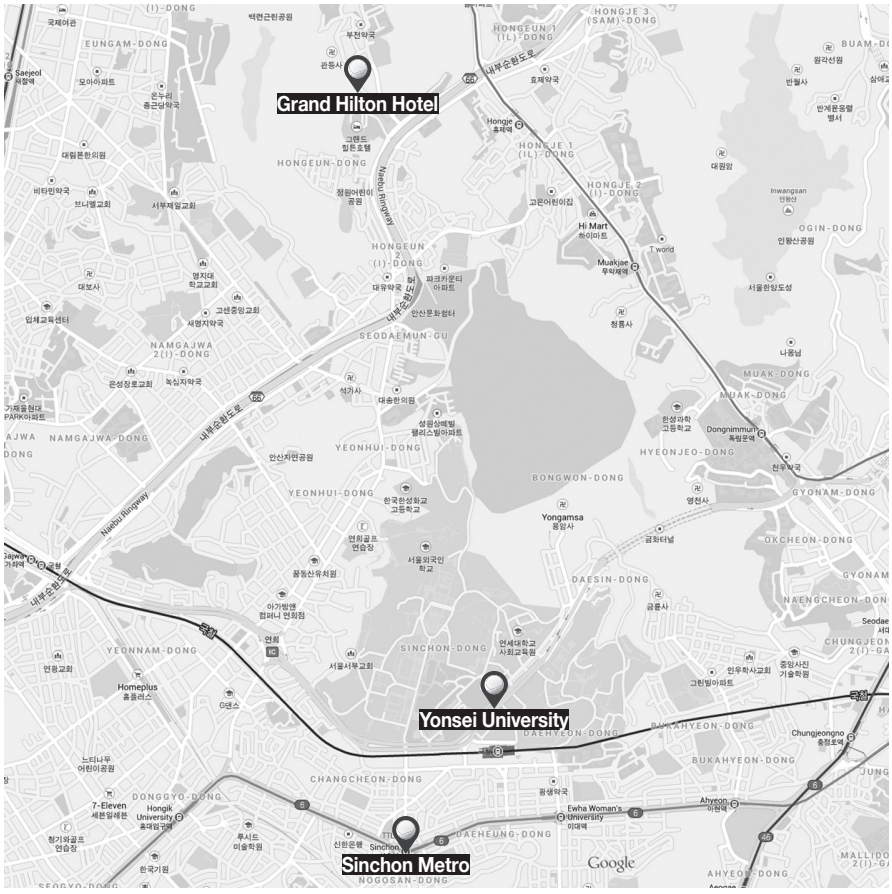
Meeting Room 2F



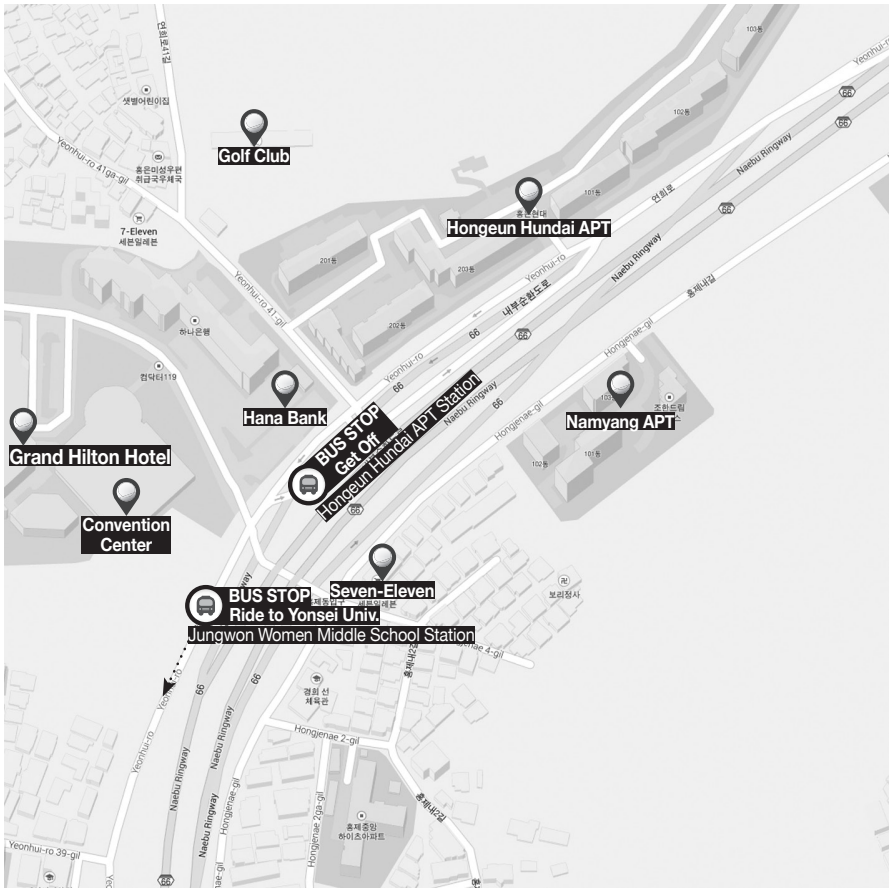
ICMPC 13 - APSCOM 5

August 4 - 8, 2014
Yonsei University
Seoul, South Korea

MAP



Grand Hilton Hotel & Yonsei University Location



Grand Hilton Hotel to Yonsei University

567, 153, 7720

Ride to.
Jungwon women's middle school (13-160)

Get off.
Yonsei University (13-115)

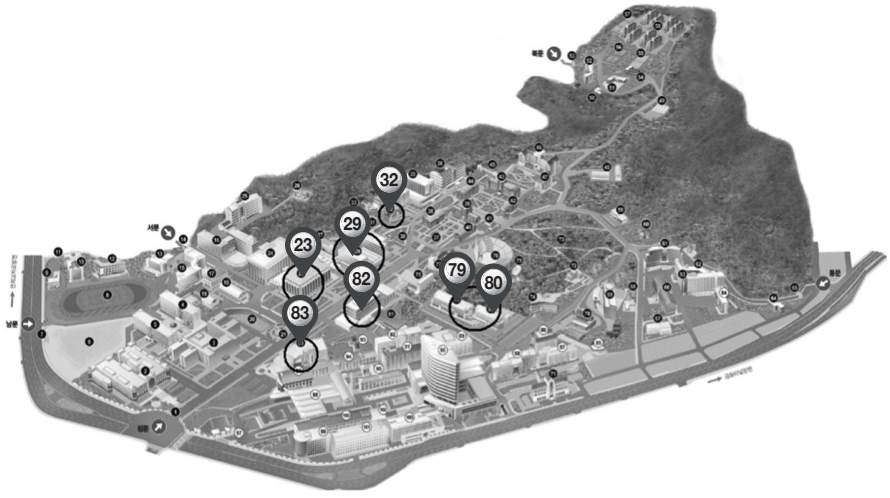
Yonsei University to Grand Hilton Hotel

567, 7720

Ride to.
Yonsei University (13-115)

Get off.
Hongeun Hyundai Apartment (13-162)

Guide



Yonsei University Campus Map

㉓ Main Library

㉙ Baekyang Hall

㉜ hankyung Hall
Today's Menu
6000won

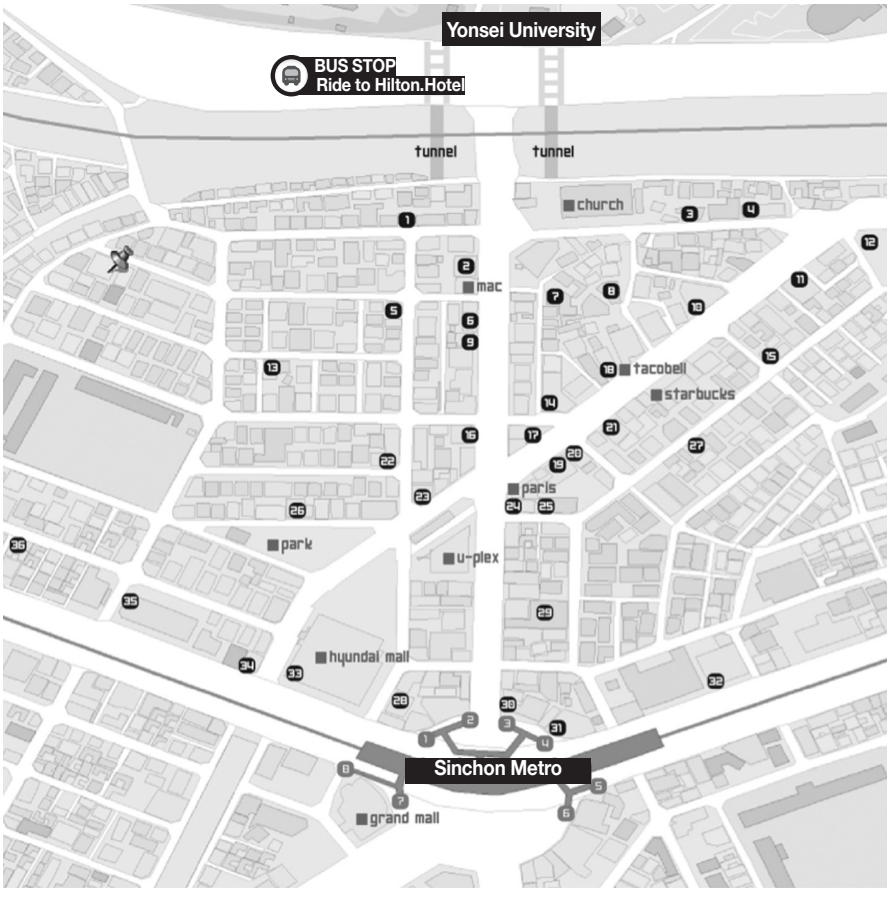
㉜ A Student Union Building
gorulsam
Pizza / Spaghetti / Gratin / Ramen
3000~5000won

Fre-fre
Korean / Japanese / Chinese / Western
3000~5000won

Basement Cafeteria
Soban / International / Noodle / Today's menu
500~3500won

㉙ ㉚ College of music

㉝ Centennial Hall



Sinchon Food Map

•Korean Restaurants

❽ ⓫ ⓭ ⓮ ⓴ ㉑ ㉓

•Chinese Restaurants

❼ ㉕

•Japanese Restaurants

❶ ❻ ㉘ ㉙

•Western Restaurants

❹ ❿ ⓬ ⓮ ⓯ ⓰ ⓳ ㉒

•Fast food

㉚ ㉛ ㉞ ⓮

•The others

❷ ❺ ❾ ⓱ ⓲ ㉔ ㉖ ㉜ ㉟ ㊱

Restaurants around the Venue

1. Yoi sushi /2nd floor **02-322-3455**
sushi 2p / 1000won

2. Loving hut *(for vegetarian)* **02-333-8088**
began burger 3500won

3. The Frypan **02-393-7707**
Chicken Tender without bones 16800won

4. Outback /2nd floor **02-3147-1871**
toowoomba pasta 19900won
buffalo beef ribs 36500won
new york strip 33500won

5. Noodle box **02-332-8952**
pad tai 7300won

6. Pungeol /2nd floo **02-3144-4981**
Okonomiyaki 12000won

7. Boksunggak **02-364-1522**
black bean pasta *(jjajangmyeon)* 4500won
pork fried sweet and sour sauce 13000won

8. Mister sin Brazier **02-333-8592**
beef galbi 19000won

9. Phobay /2nd floor **02-3144-0987**
Rice noodle 8000won

10. Team1994 /2nd floor **02-3147-0907**
carbonara 11500won
tenderioin steak 25800won

11. Sinsun sullungtang **02-393-0040**
sullungtang 7000won

12. Oktoberfest /3rd floor **02-3481-8887**
schweinehaxen *(Roasted Pork Knuckle)* 28,000won

13. Pig's day **02-337-3557**
Pork belly 4500won

14. Heong-Jae Galbi **02-365-0001**
galbitang 8000won
bulgogi 10,000~20,000won

15. CM's BOX /2nd floor
carbonara 8900won
pizza 10900won

16. Pizza Hut **02-335-5199**

17. Easy India *(Indian curry & naan)* **/3rd floor**
dinner set 16000won~

18. Elkin the black /3rd floor **02-313-0246**
carbonara*(for 2person)* 19800won

19. chosunok
Sullungtang 6500won

20. Byeokje Galbi **02-392-7892**
galbitang 19000won
bulgogi 39000won

21. Seoga & cook /2nd floor **02-324-5004**
tomato spaghetti*(for 2person)* 19800won

22. The best anchovy noodles
anchovy noodles 3000won

23. Paris Baguette **02-364-8202**

24. Manlihyang **02-393-5863**
black bean pasta *(jjajangmyeon)* 5000won
pork fried sweet and sour sauce 15000won

25. café bene **02-335-6232**

26. A to Z **02-313-7523**
homemade cheese cake 4500won
cheese panini 8500won

27. Marugamejaemyun **02-3144-3967**
Japanese noodle 6200won

28. Kauki **02-313-5505**
Rice noodle 8000won

29. Mcdonalds **02-3147-0677**

30. Burger King **02-393-5450**

31. On The Border /2nd floor **02-324-0682**
beef burrito 17500won
the ultimate fajita 33500won

32. VIPS /U-Plex 11th floor **02-3141-1997**
salad bar 27900won
ribeye steak 41400won

33. KFC **02-325-2477**

34. Ashley /B1 floor **02-326-5433**
salad bar 19900won
New York stone steak 14900won

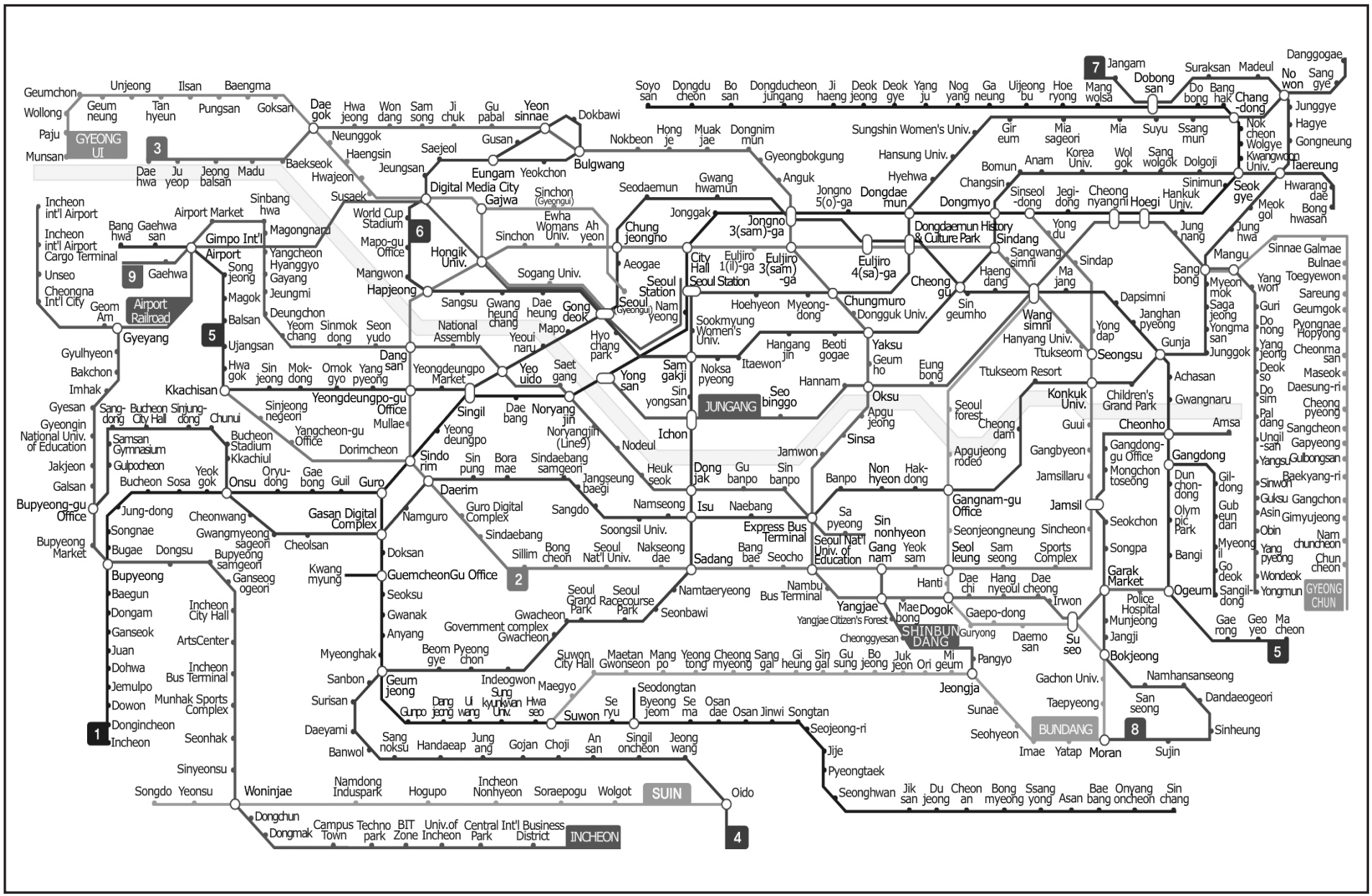
35. Mibundang **02-3141-0807**
Rice noodle 6500won

Getting Internet Access On-Campus | ID : yonsei_open / PW : none

Wireless internet access is available in the Baekyang hall. Just set up the wireless internet and install the access program by clicking 'Yonsei-info.' Once the program is successfully installed, double click the 'Yonsei wireless' icon. Input below ID. You may now have access to the internet without password. Wireless internet can be accessed by your laptop computer, Iphone, Android phone, Macbook, and Blackberry.

for Wi - Fi
FREE ZONE

Seoul Metropolitan Area Subway Network



Hosted by:

Yonsei Music Research Institute
Korean Society for Music Perception and Cognition
Korean Society for Music Theory
International Conference on Music Perception and Cognition
Asia-Pacific Society for Cognitive Sciences of Music

Sponsored by:

National Research Foundation of Korea
DaehanSteel
Korea Tourism Organization
Society for Education, Music and Psychology Research





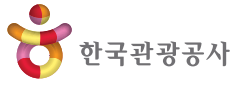














"This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government."