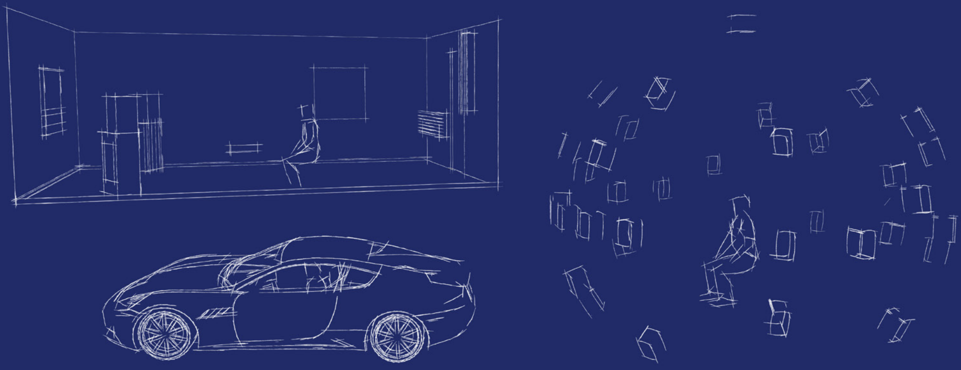
Perception of Reverberation in Domestic and Automotive Environments

Kaplanis, Neofytos

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Kaplanis, N. (2017). *Perception of Reverberation in Domestic and Automotive Environments*. Aalborg Universitetsforlag.



PERCEPTION OF REVERBERATION IN DOMESTIC AND AUTOMOTIVE ENVIRONMENTS

BY
NEO KAPLANIS

DISSERTATION SUBMITTED 2017

AALBORG UNIVERSITY
DENMARK

Perception of Reverberation in Domestic and Automotive Environments

Ph.D. Dissertation

Neo Kaplanis

Dissertation submitted December 29, 2016

Dissertation submitted:	December 29, 2016
PhD supervisor:	Prof. Søren Bech Aalborg University, and Bang & Olufsen a/s
Assistant Co-supervisors:	Assoc. Prof. Toon van Waterschoot KU Leuven, Belgium
	Prof. Søren Holdt Jensen Aalborg University
PhD committee:	Professor Stefania Serafin (chairman) Aalborg University
	Professor Steven van de Par Carl-von-Ossietzky University in Oldenburg
	Acoustic Research Engineer Rozenn Nicol Orange Labs
PhD Series:	Technical Faculty of IT and Design, Aalborg University

ISSN (online): 2446-1628
ISBN (online): 978-87-7112-869-7

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2017

The research leading to these results has received funding from the European Union's Seventh Framework Program under grant agreement no. ITN-GA- 2012-316969.

About the Author

Neo Kaplanis

Neo Kaplanis is currently a technology specialist, in sound design and research, at Bang & Olufsen a/s, Denmark. He holds a BMus Tonmeister^{TM} in Music and Sound Recording from the University of Surrey (UK) and a MSc in Music, Mind, and Brain from Goldsmiths College, University of London (UK). Neo was a visiting researcher at the Department of Media Technology at Aalto University, Finland, and at the Department of Electrical Engineering at KU Leuven, Belgium. Between 2013–2016 he was a Marie Skłodowska-Curie research fellow in DREAMS–Initial-Training-Network, residing at Bang & Olufsen a/s and supported by Aalborg University, Denmark. Previously he has been an R&D acoustics engineer at HARMAN automotive and an auditory neuroscience researcher at University of London. Over the past decade he followed his love for music by recording several classical orchestras and jazz ensembles, engineering live sound in concerts, and playing the bass.

Abstract

The central topic of this thesis is *Reverberation.* Reverberation is used as a global term to describe a series of physical and perceptual phenomena that occur in enclosed environments and relate to the acoustical interaction between a sound source and the enclosure.

This work focuses on the effects of reverberation that are likely to occur within common listening environments, such as car cabins and ordinary residential listening rooms. In the first study, a number of acoustical fields was captured in a physically modified car cabin and evaluated by expert listeners in a laboratory, using a spatial reproduction system. In the second study, nine acoustical conditions from four ordinary listening rooms were perceptually evaluated by experienced listeners. The results indicated the importance of decay times in these types of enclosures, even in these theoretically short and non-dominant quantities. It was shown that a number of perceived attributes were evoked by the alterations of the fields both within the same enclosure as well as between different ones.

The studies made use of a novel assessment framework, which forms a significant part of this work. The proposed framework overcomes previously identified challenges in perceptual evaluation of room acoustics, relating to acquisition and presentation of the acoustical fields, as well as the perceptual evaluation of such complex sound stimuli. It was shown that this framework was able to decompose the phenomena that underline the perceived sensations across assessors. The related multivariate analysis techniques employed the conjoint interpretation of both the physical and perceptual properties of the fields in a factorial space and effectively enabled the direct investigation of their relationships.

Overall the work described in this thesis contributes to: (1) understanding the perceptual effects imposed in the reproduced sound within automotive and residential enclosures, and (2) the design and implementation of a perceptual assessment protocol for evaluating room acoustics.

The thesis contains two parts. In the first, the background and rationale of the research project are presented. The second part includes four articles that describe in detail the research undertaken.

Resumé

Det centrale emne for denne ph.d-afhandling er efterklang (eng. reverberation). Focus er på rum i hjemmet (eng. domestic rooms) og bilkabiner (eng. car cabins). Der er ligeledes fokuseret på arbejde med såkaldte rum-i-rum senarier (eng. room-in-room scenarious. Ni akustiske senarier fra fire ordinære lytterum er evalueret perceptuelt af erfarne lyttere (eksperter). Resultater indikerer, at betydningen af tidsforsinkelser er utrolig vigtige for den type af rum, der studeres. Et antal af egenskaber beskriver forholdene i de enkelte rum og imellem rummene. En betydelig del af forskningsarbejdet bidrager til (i) forståelsen af implementeringen og (ii) perceptuelle vurderingsprotokoller af de rumakustiske egenskaber. Den foreslåede metode viser sig, at løse mange af de tidligere identificerede udfordringer ved perceptuel vurdering af rumakustik. Såkaldt multivariat analyse muliggør en forbundet fortolkning af de fysiske og perceptuelle egenskaber. Ph.d-afhandlingen indeholder to dele: I den første del præsenteres baggrund og rationale for forskningsprojektet. I den anden del præsenteres fire forskningsartikler som beskriver den udførte forskning.

Contents

Abbreviations

AHC	Agglomerative Hierarchical Clustering
BCS	Binaural Car Scanning
BRIR	Binaural Room Impulse Response
BRS	Binaural Room Scanning
C50/80	Clarity Index 50/80 ms
DA	Descriptive Analysis
DOA	Direction Of Arrival
DRR	Direct to Reverb Ratio
EDT	Early Decay Time
FM	Filter Model
FP	Flash Profile
HRTF	Head-Related Transfer Function
IR	Impulse Response
JND	Just Noticeable Difference
MFA	Multiple Factor Analysis
PCA	Principal Component Analysis
RIR	Room Impulse Response
RT	Reverberation Time
SA	Sensory Analysis
SDM	Spatial Decomposition Method

List of Publications

Thesis Title:	Perception of Reverberation in Domestic and Automotive Environments
Ph.D. Student:	Neo Kaplanis
University Supervisors:	Prof. Søren Bech, Aalborg University Assoc. Prof. Toon van Waterschoot, KU Leuven, Belgium Prof. Søren Holdt Jensen, Aalborg University

This thesis is based on the following publications:

[A] **N. Kaplanis**, S. Bech, S.H. Jensen, T. van Waterschoot, "Perception of Reverberation in Small Rooms: A Literature Study", *Proc. 55th International Conference of Audio Engineering Society on Spatial Audio*, Helsinki, Finland, August 2014.

[B] **N. Kaplanis**, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S.H. Jensen, "A Rapid Sensory Analysis Method for Perceptual Assessment of Automotive Audio", *Journal of Audio Engineering Society*, vol. 65, no. 1/2, 2017. Accepted.

[C] **N. Kaplanis**, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S.H. Jensen, "Perceptual Aspects of Reproduced Sound in Car Cabins", *Journal of Acoustical Society of America*, vol. , no. . Accepted.

[D] **N. Kaplanis**, S. Bech, T. Lokki, T. van Waterschoot, and S.H. Jensen, "On the Perception and Preference of Reproduced Sound in Ordinary Listening Rooms", *Journal of Acoustical Society of America*, vol. , no. . In peer-review.

The work of this thesis also relates to the publications below, reproduced in the appendix:

[E] S. Tervo, J. Pätynen, **N. Kaplanis**, M. Lydolf, S.Bech, T. Lokki, "Spatial Analysis and Synthesis of Car Audio System and Car Cabin Acoustics with a Compact Microphone Array", *Journal of Audio Engineering Society*, vol. 63, no. 11, pp. 914-925 , 2016.*

[F] **N. Kaplanis**, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S.H. Jensen, "A method for Perceptual Assessment of Automotive Audio Systems", *Proceedings 60th International Conference of Audio Engineering Society on Dereverberation of Audio, Music, and Speech (DREAMS)*, Leuven, Belgium, February 2016.

The work of this thesis relates to other collaborative publications:

- G. Vairetti, E. De Sena, T van Waterschoot, M. Moonen, M. Catrysse, **N. Kaplanis**, S. H. Jensen, "A physically motivated parametric model for compact representation of room impulse responses based on orthonormal basis functions", *Proceedings of the 10th European Congress and Exposition on Noise Control Engineering – EURONOISE*, Maastricht, The Netherlands, 2015.*
- G. Vairetti, **N. Kaplanis**, E. De Sena, S.H. Jensen, S. Bech, M. Moonen, T. van Waterschoot, "The Subwoofer Room Impulse Response database (SUBRIR)", *Journal of Audio Engineering Society*, Submitted 2017.*
- E. De Sena, **N. Kaplanis**, P. Naylor, T. van Waterschoot, "Large-scale auralised sound localisation experiment", *Proceedings 60th International Conference of Audio Engineering Society on Dereverberation of Audio, Music, and Speech (DREAMS)*, Leuven, Belgium, February 2016.*

* the author of this PhD thesis was only a contributor to this work.

The work leading to Papers [B], [C], and [D], included a series of investigations in several domains relating to acquisition, analysis, and presentation of sound fields of car cabins and domestic listening rooms. A database of spatial and binaural room impulse responses was built, and used to apply *Spatial Decomposition Method* (SDM) in near-field reproduction within car cabins. This work is included in appendices, labeled as Paper [E] and Paper [F].

Similar investigations were conducted within ordinary listening environments. There, the low-frequency distribution in rooms was investigated, leading to a collaborative project within DREAMS network, which addressed the calculation of reverberation time at low frequency bands, summarized in Vairetti et al., 2015 & Vairetti et al., 2017. The related measurements were also made publicly available. As part of DREAMS dissemination work, collaborative work with E. De Sena, P. Naylor, and T. van Waterschoot led to publication of De Sena et al. 2016, regarding localization of sound sources in reverberant sound fields.

The author contributed to a collaborative project between Technical University of Denmark, Widex a/s, and Bang & Olufsen a/s, titled “Control of spatial impression by means of apparent source width”, funded by the Danish Sound Innovation Network. The project looked into the perception of source width in normal and hearing-impaired listeners when the sound field is reproduced over loudspeakers in a typical room.

* * *

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the submitted or published scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty. The thesis is not in its present form acceptable for open publication but only in limited and closed circulation as copyright may not be ensured.

Preface

When addressing complex and exploratory research questions, such as understanding the auditory human perception in enclosures, it is important to define and understand the rationale and motivation of the investigation. That is, the way the work undertaken will enhance a general research purpose, via a restricted and application-specific experimental protocol. This is illustrated in the following paragraph by a top-down approach, starting from the *Why* the research was undertaken, leading to the *What* has been investigated.

The work described here is a part of the *De-reverberation and Reverberation of Audio, Music, and Speech* (DREAMS) Initial Training Network, a European funded project working in the general framework of Reverberation and De-reverberation. DREAMS' research motivation focuses on understanding reverberation and its properties, in an attempt to better control its influence in sound fields. That includes the improvement of speech intelligibility for the hearing impaired in reverberant environments, as well as the enhancement of human experience in an entertaining scenario, such as listening to music at home. It was the project's purpose, its *Why*, to contribute to the improvement of the perceived sound experience in typical listening environments.

Understanding both the physical and perceptual aspects of reverberation would allow the development of perceptually-relevant algorithms. That is, algorithms that could improve the human experience in reverberant spaces, by employing advanced signal processing techniques to better compensate for the acoustical degradation. That is the study's *How*, the approaches and methods used to achieve this.

The current thesis, contributes to the *What*. That is the objective and research scope of the studies that follow. The work in this thesis aims to understand, identify, and quantify the perceptual effects evoked in reverberant fields, and attempts to depict the perceptual importance of certain properties of a sound field. Understanding the perceptual aspects of these spaces and their relation to the physical characteristics will enable the development of perceptually-relevant algorithms allowing the perceived aural experience in such spaces to be room-independent; the project's *How*. By implementing such protocols the human experience in such fields could be improved, thus, serving the project's high level objective, its *Why*. Figure 1 summarizes the rationale of this study, and its link to the general purpose of the DREAMS objectives.

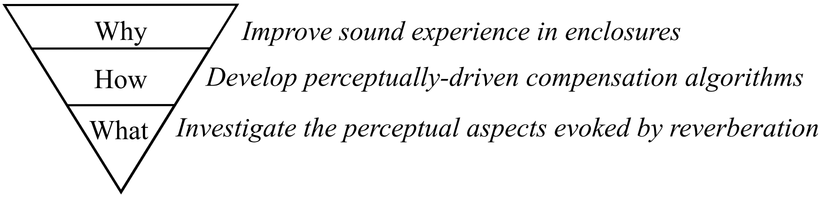


Fig. 1: The rationale and motivation of this thesis relating to DREAMS' research objectives. This illustration follows Simon Sinek's "Start with Why".

* * *

This thesis is submitted to the Doctoral School of Engineering and Science at Aalborg University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The work was carried out in the period from August 15, 2013 at Bang & Olufsen as well as at the Department of Electronic Systems at Aalborg University. The research leading to these results has received funding from the European Union's Seventh Framework Program under grant agreement no. ITN-GA-2012-316969.

Neo Kaplanis
Bang & Olufsen / Aalborg University, December 29, 2016

Acknowledgements

"One doesn't discover new lands without consenting to lose sight of the shore for a very long time" – André Gide

When I first committed to this project, there was no doubt that the period will be marked with new experiences, several mistakes, and inspiring learning trips. It was expected to require a lot of effort, long times in the lab, large cups of coffee, and endless motivation. That was not far off.

Yet, the most important of all was indeed the amazing bunch of people that kept me going forward. First, my supervisor Prof. Søren Bech, who I truly thank for our multiple and thoughtful discussions, his personal interest in the project, his superb leadership skills, his support, and trust to myself from day one to date. My academic shields, Prof. Søren Holdt Jensen and Prof. Toon van Waterschoot for their numerous contributions, their inspiring characters, and their support throughout this project. I am thankful that we completed this journey together, and proud that I had you by my side along the way.

Further, I would like to thank Prof. Tapio Lokki and his team, Dr. Jukka Pätynen, and Dr. Sakari Tervo. Although not an official part of the project, they supported my research interests at the highest level and became an integral part of this work. Thank you guys for hosting me at your lab, sharing your experiences, time, and skills for this project. It has been a pleasure to work with you all along.

I am thankful to the colleagues I had over these three years. Martin Møller and Martin Olsen, for their useful inputs to my work, their challenging questions, and their inspiring passion in the field of acoustics. Morten Lydolf, Patrick Hegarty, and Adrian Celestinos for their initial support in conducting acoustic measurements, their awesome MATLAB scripts that they shared with me, and their amazing characters that enlightened the office on daily basis. Jesper Kjær Nielsen, Peter Websdell, Samuel Moulin, Jussi Rämö, and Sven Shepstone for our daily discussions and the instantaneous ideas that popped up, some of which ended up in this work. That includes my director, Lars Jørgensen, who truly supported my last efforts in this work.

A huge thank you to the people who surrounded me when I was not working (these very few minutes) or joined my working habits (the most often), Photini (i.e.

Dr. Vrikki), Maria, George, Stavros, and Christos. Your love, support, and belief took this steps further, if not levels.

Last but not least, my other half, Dr. Maria Spyrou, who I have been lucky enough to have by my side for years, even when tough times were ahead. Thank you for following my unconventional habitual life and joining me in various cities (and countries), and kept a fun and passionate character when it was really needed. You are awesome.

Neo, 2016

"*Living is a form of not being sure, not knowing what next or how. The moment you know how, you begin to die a little. The artist never entirely knows. We guess. We may be wrong, but we take leap after leap in the dark.*" – Agnes De Mille.

Part I

Introduction

Sound in Enclosures

1 Introduction

Typically, the sound that arrives at our ears is a distorted variant of the sound emitted by a source. The perceived sound is a mixture of the original sound, as emitted by the sound source, and a combination of absorbed, diffused, and modified copies of it; commonly referred to as acoustic *reflections.* This occurs as the emitted sound interacts with the geometrical characteristics, boundaries, and objects within the surrounding environment as it propagates through the space. The effect is particularly prominent when humans experience sound in large spaces, such as concert halls. This phenomenon is normally attributed to *Reverberation.* Reverberation is a global descriptor and forms a common lexicon that uniquely expresses both the perceptual and the acoustical properties of an enclosure [1]. In this thesis, reverberation is used as the holistic description of the phenomena observed when sound propagates in enclosures.

Reverberation does not occur only in performance spaces but it is a phenomenon we encounter constantly. It is unlikely that we experience a reflection-free environment in daily life. Even when we experience sound outdoors, a strong reflection from the ground can be observed which alters the perceived sound experience [2]. Arguably this effect is not as prominent to the human ear as the perceived sound inside a large space. Yet, both experiences operate on the same acoustical and psychological principles.

The perceptual influence of reverberation to the human listener drives both desired and undesired effects. In a concert hall, the interaction between the sound produced by the orchestra and the acoustical characteristics of the enclosure enhances the performance and reveals a perceptually pleasing result to the human listener. However, when speech communication is required in such spaces those effects reduce speech intelligibility, resulting in an uncomfortable experience. It is unequivocally evident that the phenomenon of reverberation is driven by the physical and acoustical characteristics of the environment, demonstrating the importance of understanding these and identifying their effects on human perception.

In a series of studies [3] during early 1920's, Physicist Wallace Clement

Sabine investigated the relationship between the physical and perceptual properties of spaces; effectively founding the scientific field of architectural acoustics. A plethora of investigations in that domain have been conducted since, aiming to study the effects of reverberation in multiple environments including concert halls [4–7], opera houses [8], historical theaters, and auditoria [9]. The rationale and motivation behind these studies followed a common scope. That is, to understand the properties of the reverberant sound field and identify the links between the geometrical architecture, the acoustical properties, and the perceived aural qualities.

The effects of reverberation in geometrically smaller environments that we are likely to experience in everyday life are not well understood, and the research in the domain is limited [10]. The geometrical divergence between the purpose-made performance spaces and these enclosures has significant effects in both the physical characterization of the fields as well as the perceived sensations [11]. Moreover, these spaces are typically built to serve multiple purposes and uses, for example speech communication, i.e., classrooms or living dwellings, i.e., bedrooms. In this research the focus is on typical scenarios that are likely to occur in ordinary listening settings. There, the sound source is typically a sound reproduction system, i.e., a loudspeaker, encompassing its own characteristics, which are known to influence both the timbral and the spatial characteristics of the aural experience [12, 13].

It is evident that every environment we encounter has a distinct and unique sonic identity that we experience as a set of sensations. This defines the general sound of a church, or the unique properties of a specific church we experienced before, a hall, a living room, or a car. Yet, it is still unknown which acoustical properties evoke certain sensations and how these relationships operate.

Here, we aim to identify the influence of acoustical characteristics of everyday environments on human perception, and seek to characterize and quantify their relationships. The motivation behind this thesis follows the limited understanding of the perceptual effects of acoustical properties in enclosures e.g., reverberation, as encountered in listening environments we experience every day. The investigations in this work focus on the two most common listening environments [14]: the automotive car cabins and ordinary residential listening rooms.

1.1 Research Questions

Before the related background and research findings of this thesis are presented, it is important to define the questions that directly address the research interests of this work. Namely, our attempts to better understand the perceptual effects of acoustical properties of commonly encountered listening spaces, such as the automotive and domestic environments. This thesis directly addresses three research questions:

[RQ1] *What are the most relevant attributes that are likely to characterize the perceived sound experience, i.e. reverberation, in domestic rooms and car cabins?*

[RQ2] *What physical properties of the sound field are likely to affect the human experience in these environments?*

[RQ3] *What is the relationship between the identified perceptual attributes and the physical properties in these fields?*

These research questions defined the direction of the work presented in this thesis. During the attempts to address them, other questions were stipulated which further expanded our scope. These are further discussed in the following sections.

1.2 Thesis Structure

This thesis comprises of two parts. Part I presents the relevant background, including the rationale of this investigation and the major findings. Part II includes four articles, depicting the outcome of this work, and describing the studies in detail.

In the following sections of Part I, the thesis provides a brief introduction to the principles of sound propagation in enclosures relating to the physical characterization of the phenomenon in Section 2.1, and the effects to human perception in Sec. 2.3. This is followed by presenting the fundamental components of perceptual assessment protocols in Sec. 3, where the state-of-the-art protocols are critically assessed, focusing on their limitations and the requirements of evaluating the acoustics in domestic and automotive environments, summarized in Sec. 4. The overview of this thesis is given in Sec. 5, including the contributions and dissemination of the work presented here, and finally the conclusions, limitations, and future work are discussed in Sec. 6.

2 Background

2.1 Basic Principles of Sound Propagation in Enclosed Spaces

When a sound is produced in an enclosure, the sound waves propagate in the finite space at a finite speed, and they arrive at a receiver, e.g., the human listener, after a temporal delay. During this transmission, the waves interact with the geometrical properties of the enclosure introducing multiple reflections, which cluster together at narrow time intervals. As a consequence the received sound is a modified version of the produced sound. This is driven by the physical characteristics of the *enclosure*, the properties of the excitation *source*, and the characteristics of the emitted sound, i.e. the *signal*, as graphically summarized in Fig. 2.

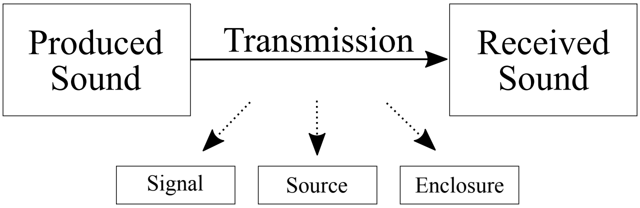


Fig. 2: Basic principles of sound propagation in enclosures. The produced sound is being transmitted to a receiver, e.g., the human ear. The reception of this sound depends on the transmission process, which relates to the characteristics of the signal, the source, and the enclosure.

As mentioned above, the result of this acoustical process is typically referred to as reverberation. In order to physically describe reverberation a simplification is commonly followed, which resembles the series of events that occur when an impulsive sound is produced in an enclosed space by an omnidirectional source, normally expressed from the receiver's point of view. This simplification is realized as an acoustical *Room Impulse Response* (RIR), describing the temporal characteristics of the sound field and energy. A typical RIR is shown in Fig. 3.

The first sound that arrives at the receiver describes the sound that propagated between the source and the receiver following the shortest distance, i.e. in a straight and direct path, and it is referred to as the *direct sound*. Later, *early reflections* arrive at the receiver, which is the result of the interaction between the propagating wave and the nearby objects or boundaries. The intensity of these reflected sound waves is typically reduced due to the absorptive nature of room's boundaries. The reflected energy continues to spread in the finite space of the enclosure, interacting further with these surfaces. This interaction decreases their intensity over time and results in a well-known exponential decay pattern that typically describes the behavior of *late reflections*.

The temporal division of the field in the two reflection zones, that is, the earlier, distinct, and strong reflection patterns, and the later, denser, and decaying reflections formed the fundamental principles behind the physical and perceptual understanding of sound in enclosures, and reverberation *per se*. Fundamental work in architectural acoustics [3] has shown that the exponential decaying nature of the later reflections is largely similar in various positions of a room. Sabine's observations led to a metric that defined the time needed for a sound intensity to decrease by factor of million, i.e., 60 dB. The reverberation time, *Reverberation Time –60dB* (RT_{60}), has been followed since describing the decay time of a sound field in an enclosure. Following this stream of re-

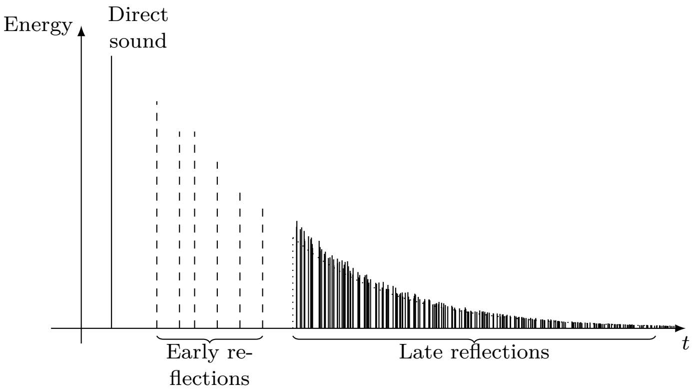


Fig. 3: Simplified Room Impulse Response (RIR). This resembles the simplified timeline of events in a reverberant sound field.

search, reverberation has been mainly considered as a temporal phenomenon that describes the prolonging sound in enclosures, highlighting the nature of the surrounding surfaces [9].

2.2 Decomposing Perception - The Filter Model

The human responses evoked by the physical properties of the sound field are fundamental to establish and quantify the relationships between the perceived sensations and measurable physical quantities. It is evident in the literature that the physical and perceptual features of reverberation are indivisible. In order to better understand the effects of this phenomenon, one needs to address the two domains conjointly.

Yet, the non-linear and highly adaptive auditory system humans posses [15], makes the direct investigation of physical properties of audio signals and the perceived sensations a complicated task [16]. Moreover, when the measuring instrument is the human perception, it is likely that the observations would follow inherent cognitive biases, such as own references, beliefs, and sentiments. It is therefore critical, that the assessment protocols followed in such evaluations address these issues, so that robust and repeatable results are observed.

Based on these principles, the *Filter Model* (FM) [16–18] provides a framework that differentiates the qualities and properties of a sound stimulus, in different domains. In general, the sound can be described in terms of its *physical* characteristics, using instrumental means, and in terms of the *perceptual* characteristics using human assessors as the measuring device. The FM is shown in Fig. 4.

The model's *Physical Domain*, addresses the characteristics of the field based on instrumental metrics, i.e. mathematical and physical descriptions

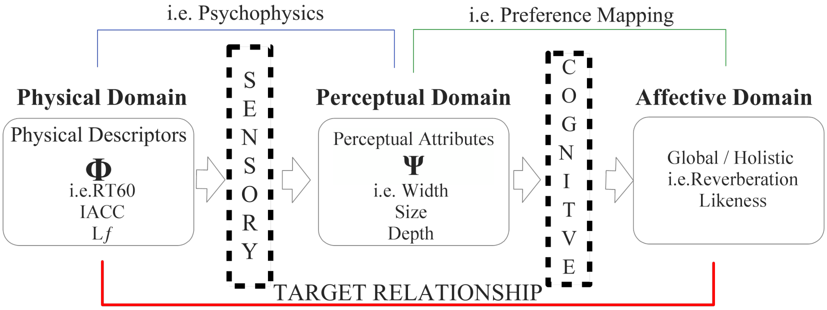


Fig. 4: The Filter Model, comprising three domains, physical, perceptual, and affective separated by a sensory and a cognitive filter respectively. Reproduced from Paper [A].

of the sound stimulus, such as sound pressure levels, frequency-based metrics and reverberation time. When the characterization of the sound is based on measurements using human assessors, the model defines two separate domains, namely, the *Perceptual* and *Affective* domains. This distinction follows the properties of the human auditory system, illustrated by the two main filters of the model, the *Sensory* and *Cognitive* filters. The sensory filter encompasses low-level psychoacoustical and physiological mechanisms of the auditory system, such as frequency selectivity of the cochlea and auditory masking, which are assumed to be common across humans. The cognitive filter relates to mechanisms at higher levels of the auditory processing, such as expectations, mood, and memory. These processes are likely to evoke responses that are individual-based, uncontrolled, and biased to one's background.

Therefore, in the *Perceptual* domain one could seek for discriminative, uni-dimensional, and defined features of the sound, known as *perceptual attributes*. That is, the quantitative measurements of the sensations evoked from a sound stimulus, that could be used to quantify and characterize the perceived aural experience. In the *Affective* domain the sound is typically assessed as a holistic experience, relating to general quality, such as likeness, and preference.

In this work, the basic principles of FM are followed. The abstractions of physical, perceptual, and affective domains are further referred to the FM's definitions.

2.3 Reverberation & Perception

For years, the relationship between RT_{60} and the perceived *reverberance*[1] dominated the design of 'acoustically prime' performance spaces, following optimum RT_{60} figures in the range of 1.8 to 2.2 seconds [19]. It was however realized, that the decay time and the perceived prolonging sound, could not fully describe

---

[1] the the sense of a prolonging sound that decreases over time [10]; a prominent sensation when auditioning in geometrically large spaces.

the properties of reverberation. That is, two spaces with identical reverberation times, were unlikely to be characterized as physically and perceptually identical.

Although reverberation, can be perceived as a holistic entity, describing the global behavior of an acoustic environment, i.e., the perceived *Reverberance*, it evokes a multitude of other perceptual sensations [10], all of which contribute to the sonic identity of the space. Investigating these phenomena in the late 40s, Lothar Cremer suggested [20] that the properties and patterns of both early and late reflections had a significant effect in the acoustical qualities of the space. Follow up studies [21, 22] identified the significance of the first arriving reflections and their relationships.

These observations steered the acoustical research towards the objective and physical characterization of the reverberant fields. The research stream introduced a number of the physical and mathematical depictions of properties of reverberation that have been now well established and standardized [23, 24]. These metrics include RT_{60}, *Early Decay Time* (EDT), *Sound Strength* (G), *Clarity Index* ($\mathrm{C}_{50/80}$), *Interaural Cross-Correlation* (IACC) and many others [23, 25, 26]. Although a number of physical descriptors exist and are used widely, their perceptual relevance, that is, the extent to which these metrics could explain the perceived experience, is highly challenged [27, 28].

These findings evoked further the central question in architectural acoustics research: *"What defines an environment to be perceived as a good concert hall?"*. That is, *"Which properties of the acoustical field evoke certain perceived sensations, and what defines the aural qualities and identity of a sonic environment?"*.

A large number of studies attempted to answer these research questions, primarily focusing on performance spaces and the musical enjoyment in concert halls and opera houses. Several perceptual aspects of reverberation have been identified and links to physical metrics have been proposed. For example, metrics relating to decay time of a reverberant field, RT_{60} and EDT, have been linked to the perceptual sense of *Reverberance*, and *Room Size*, whilst the relationship between the early and late reflections, measured by $\mathrm{C}_{50/80}$ have been related to the perceived *Clarity* of the sound.

A comprehensive review of these investigations has been conducted as part of this thesis, presented in Paper [A]. Through this in-depth analysis across research disciplines, several trends have been identified and reported; see Fig. 5. It was however apparent that the relationship between physical and perceptual metrics is still a field of research debate. It is well understood that reverberation includes a number of dissimilar perceptual sensations that are described by certain physical metrics. However, it is unlikely that these relationships would occur independently in real sound fields. Although several perceptual properties and physical characteristics of reverberation were identified, their relationships have been physiologically quantified in highly controlled and simulated sound fields [29]. The applicability of these relationships to naturally occurring fields is therefore not straightforward, and currently only partial and application-specific understanding of the phenomenon is apparent.

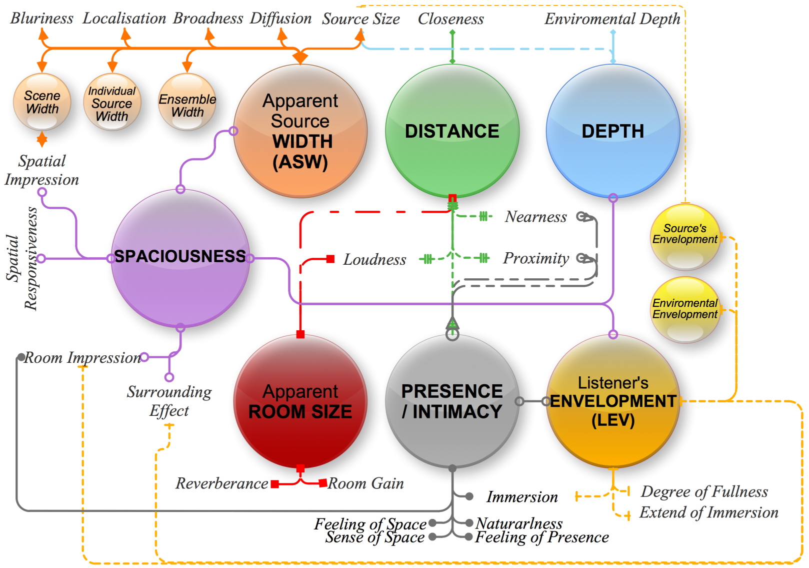


Fig. 5: Overview of the findings, indicating the major perceptual attributes and their relationships as described in reviewed literature. Reproduced from Paper [A].

The state-of-the-art in perceptual assessment of architectural acoustics focuses on the identification of co-occurring characteristics in reverberation fields by comparing the features of real spaces [5, 6, 30]. That is, the perceptual and physical properties of real environments, i.e. concert halls, are directly compared so that common trends and dissimilarities between the two are identified. Although this approach is not novel in architectural acoustics [4, 7], today's technological advancements provide a toolkit that allows researchers to overcome several shortcomings that prohibited a repeatable, robust, and controlled experiment between real-sound fields. It has also been noted that in the last decades, architectural acoustics studies have focused on investigations that conjointly quantify the perceptual and physical aspects of the fields, in contrast to the previously separate streams of research. That is a significant finding, and highlights the complex relationships between the two domains.

At this point, it will be useful to review the current practices in perceptual assessment of room acoustics. This will allow us to identify possible limitations of the state-of-the-art and stipulate further considerations for assessing room acoustics properties of domestic rooms and automotive car cabins. In the sections below, the fundamental principles of perceptual assessment of audio are presented, focusing on the requirements of assessing room acoustics properties in a typical sound reproduction scenario.

3 Perceptual Assessment of Room Acoustics

In the lack of a standardized protocol and evaluation procedures, a wealth of scientific studies attempted to perceptually evaluate the qualities and acoustical properties of several spaces, following general principles of audio assessment [16]. Two basic frameworks were typically followed to evaluate the perceptual qualities of sound in enclosures. These include the *in-situ* evaluation and the *laboratory* based protocols. The former requires the human assessor to evaluate the sound field/stimulus in natural settings, for example, a pair of loudspeakers in a listening room. The latter, follows an evaluation protocol, where the sound field/stimulus is captured in real settings or by means of computer simulation, and then recreated in neutral environments, where the listening experiment is conducted.

In the domain of concert hall acoustics, *in-situ* assessment seemed a natural choice and has been followed consistently since the early days of architectural acoustics research. In these protocols the assessor's perceived sensations where typically observed during real performances using structured questionnaires and *post-hoc* interviews [4]. The practical limitations of these methods and the need for repeatable experimental settings and procedures gave rise to laboratory-based methods. In these methods, the sound field of an orchestra in several concert halls was recreated [29] over a sound reproduction system and evaluated by human assessors in a comparative and repeatable manner.

Perceptual assessment of sound in smaller listening enclosures has been also conducted, both *in-situ* [31] and in the laboratory, using simulations of the acoustical field [32] as well as binaural measurements [33]. In these studies however, the dominant research interests related to the effects of the interaction between the source's properties, i.e., the loudspeaker, and the room. The sound reproduction has therefore dominated these research studies, where the acoustical properties of the listening enclosures were often simplified or assumed to be constant.

This work seeks to investigate the imposed effects of the acoustical properties of domestic and automotive enclosures on the perceived sound field, by identifying and quantifying the perceived sensations that relate to the acoustical features of these spaces.

When investigating the effects of room acoustics and reverberation *per se*, one needs to employ a protocol where the influencing factors are controlled and a systematic variation of the acoustical properties of the enclosure is followed. Moreover, the acquisition and presentation of the acoustical field should be experimentally repeatable, so that different assessors are exposed to identical settings and sound fields. It was therefore important to identify the influencing factors of such investigations and define a framework that would enable the perceptual assessment in a complex sound field. These considerations were initially stipulated in Paper [A], where the applicability of current methodologies to this research project was discussed. Several limitations and requirements were presented and further highlighted the need of a new experimental framework for the evaluation of acoustics in small enclosures such as domestic rooms and automotive car cabins.

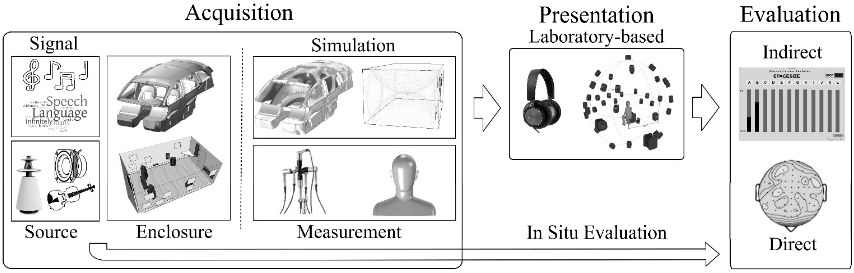


Fig. 6: Basic principles of audio evaluation, depicting the procedures followed in *in-situ* evaluations and laboratory-based methods. The figure is adapted from Paper [B] [34].

In perceptual assessment of sound in enclosures, the protocols employed typically follow the basic principles shown in Fig. 6. One could classify these in three major parts, the *acquisition* of the field, its *presentation* to human subjects, and the *evaluation* procedure where human responses are observed. In such evaluation schemes, the stimuli under investigation are the product of a signal, a source, and an enclosure, each of which encompasses its own characteristics. As it is discussed in Sec. 4.1 the properties of these factors highly influence the sound field and one is required to impose experimental controls when evaluating the properties of the enclosure.

In the sections below, the influencing factors are discussed in detail and considerations are raised following our research interests. Then the available acquisition and presentation methods are described, in Sec. 3.4, followed by the evaluation protocols in Sec. 3.7. General remarks are given in Sec. 4, where the limitations of these protocols are summarized.

3.1 The Excitation Signal

It is well understood that reverberation is both a spatial and a temporal phenomenon [1]. It is therefore critical that the signals, as well as the sound sources, excite both aspects of the fields under investigation. In order to assess spatial attributes one needs to use broad-band and realistic signals and sources, that exhibit dynamic scenes and content [35]. Yet, the temporal properties of reverberation are more prominent with impulsive signals, which led to many studies incorporating narrow-band noise bursts [10].

In concert hall research, the excitation signal was not considered as a major factor at first, as *in-situ* evaluations occurred during live orchestra performances, in consequence, there was no experimental control on the repertoire. It has been recently noted by laboratory-based studies [36] that the properties of the signals, i.e., the music piece, used to excite the hall altered the perceived sensations significantly. Similarly, standardized audio assessment protocols [37, 38] often recommend the use of a number of different signal types and musical genres [16] so that signal-based effects are highlighted.

Moreover, the perceived sound field in a residential listening environment, via loudspeakers, is likely to be comprised of two acoustical fields. That is, the acoustic response of the recorded room, where the signal was captured, as well as the acoustics of the reproduction room, where the loudspeaker reproduces that signal. This is typically referred to as *room-in-room* effect [39] and it has been shown to affect listener's spatial awareness whilst listeners are able to discriminate the two fields independently [35]. That is a major difference to concert halls assessment, where the signal is typically the natural sound of the instruments.

It seems therefore appropriate to our research interests, that the excitation signals include both less reverberant recordings, e.g., anechoic, as well as typical dynamic and well reverberated signals, e.g., music. These would result to a controlled and repeatable protocol that extends in both cases, yet, they would reveal a familiar and natural signal to the human listener.

3.2 The Sound Source

A performance hall is typically designed and built to accommodate an orchestra, which is placed at a pre-defined point, i.e. the stage, and it is experienced at a set of known listening positions, i.e. seating map. In every-day environments the source is unlikely to be an orchestra but it is rather a sound system that comprises of several loudspeakers, that may be arbitrarily placed in the room. This adds another layer of complexity when the acoustics of these types of enclosures are in question. Loudspeakers differ in many ways and it is well known that their properties and placement in the room would highly influence the perceived sound field [12, 13].

It is anticipated that the excitation of the field by a typical loudspeaker in an ordinary room and an orchestra in a hall would result in dissimilar physical and perceptual observations. For example, most standard physical quantities used to describe performance halls make use of an omnidirectional source; simulating the directional characteristics of an orchestra. This scenario is unlikely to be observed in a typical listening environment [40], thus its perceptual relevance is highly challenged. In contrast to concert halls, ordinary listening environments are geometrically smaller and include strong, distinct, and dense early reflection patterns. The properties of these reflections are highly influenced by the loudspeaker's placement and directivity characteristics, affecting the perceived *Timbre*, *Apparent Source Width*, as well as the location of the sound source [40].

This study aims to assess the effects imposed by the reproduction room. Therefore, the source and signals should be kept constant, while the properties of the enclosure are modified. The source should represent a typical reproduction system that one is likely to experience in that type of enclosure, whilst its position and main characteristics are well controlled.

In our research interests, we consider two types of enclosures, that is, the ordinary residential listening rooms and car cabins. Although in rooms a typical reproduction scenario could be kept constant and be comparatively evaluated, e.g., a stereo configuration in several standardized listening rooms [41, 42], in automotive environments such approach is not possible. Each car audio system is individually designed and it is unique in its own manner, thus comparing several car cabins, while the reproduction system is kept constant, is unlikely. These are challenges that should be addressed in the experimental designs within automotive audio evaluation.

3.3 The Enclosure

Naturally, a geometrically smaller space is expected to be characterized by shorter decay times due to the lower volume [43] they comprise. The volume, geometrical shape, and the acoustical properties of the enclosure are likely to affect the observed decay times, as well as the early reflection patterns in terms of their temporal distribution, amplitude, and density. These observations are shown in Fig. 7, where the RIR of three types of enclosures, a concert hall, a listening room, and a car cabin are contrasted.

As seen in Fig. 7, the reflection patterns are dissimilar between the three fields, both in terms of energy and temporal distribution. Standardized physical metrics [23, 24] that are used to characterize performance halls, seem unsuitable for describing the properties of smaller enclosures. For example, time-based metrics such as *Clarity Index (80 ms)* (C_{80}), assume a temporal distribution of early and late reflection patterns that is unlikely to occur in a living room, or a car cabin. The temporal and spatial distribution of the reflections in smaller rooms, may enhance specific auditory processing mechanisms and perceptual weightings, i.e., dominance of *echo suppression*[2] and *precedence effects* [44]; driven by the reflections' intensity, spectral properties, as well as their temporal

---

[2] echo suppression occurs when two coherent signals are presented to the ears from different locations with a slight delay. A human listener perceives only the first event, rather than two distinct sound events, as the human hearing system assigns priority to the directional information of the first event, effectively suppressing the second, thus, allowing faster processing in complex scenarios.

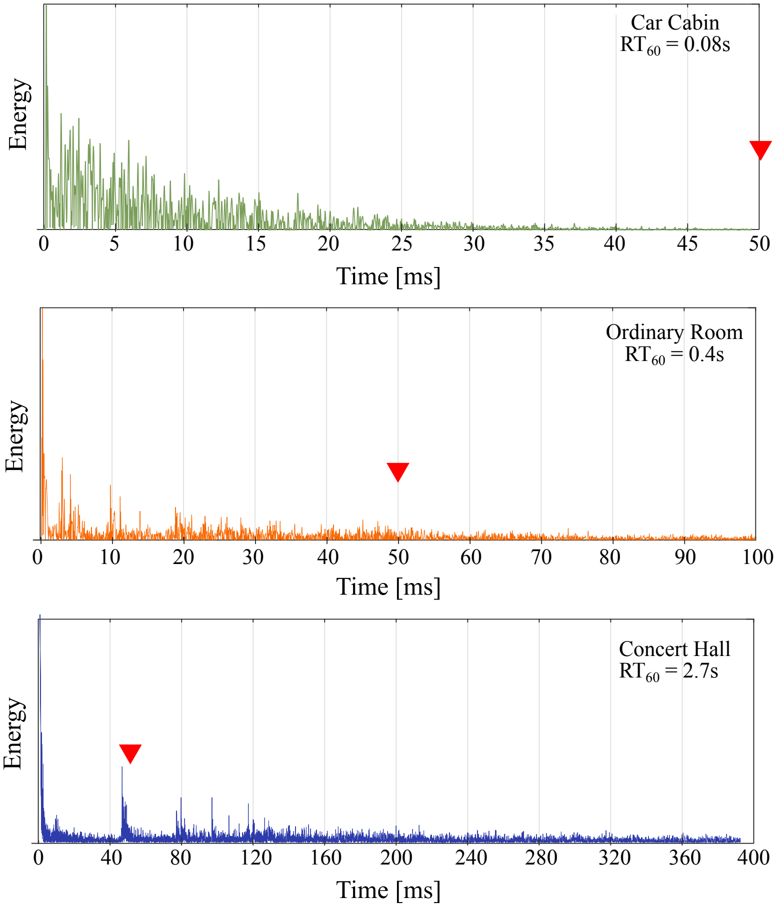


Fig. 7: Room Impulse Response measured in a car cabin, ordinary listening room, and concert hall. The responses were normalized and are partially displayed for illustration purposes, focusing on the early reflection patterns. Note that the horizontal axis follow different scaling in the three responses due to the temporal properties of the fields. The red triangle indicates the 50 ms mark.

distribution [45, 46]. Low-level psychoacoustical mechanisms may alter the perceived sensations in a dissimilar manner, relating to the field the human auditory is exposed to.

In order to study the properties of the enclosure, one needs to systematically modify their physical characteristics so that a range of sound fields is obtained, or study several enclosures of the same type, and comparatively identify commonalities. It is well known that real-life enclosures differ in a myriad of ways (size, materials, geometry). Thus, introducing a systematic variation of isolated parts of the sound field is almost impossible in practice. Both the physical and perceptual aspects of the field are unlikely to alter independently when real spaces are in question.

Nevertheless, a real room can be physically modified, by altering the absorptive properties of the boundaries' materials. This approach requires a modular room and it is laborious. Moreover, the observed changes will include several alterations of the field, rather than a single quantifiable change, e.g., *Clarity Index (50 ms)* (C_{50}) in isolation. It is however possible to achieve single alterations in the field by employing computer simulations, such as geometrical acoustics modeling [47]. The conceptual distinction that emerges between the two approaches is however important; a perceptual space identified in a controlled modification of a field, i.e., in a simulation, may not relate to a natural acoustic change that is likely to occur in real life.

These challenges would typically result in domain specific experimental designs, focusing on the contextual factors and applications of interest. This includes the method to acquire and reproduce a sound field, as well as the experimental designs to observe human responses during evaluation. There are several approaches that one could employ to address these. Below, a brief overview is given, focusing on the considerations of our research interests.

3.4 Experimental Apparatus: Acquisition & Presentation

A fundamental decision when conducting perceptual assessment of audio material is whether the signals are evaluated *in-situ*, in their intended and natural settings, or in a neutral and controlled *laboratory*. Through the literature one can identify a major trade-off between the two approaches. *In situ* evaluation provides the listener with a familiar and natural experience, thus, the evaluation of the perceived sensations relate directly to reality. It is however difficult to employ comparative protocols and systematic alterations in such evaluation schemes, leading to potential uncontrolled biases relating to both auditory and non-auditory factors. Laboratory-based methods are able to overcome these challenges, yet, their perceptual relevance and ecological validity could be questioned.

3.5 In-Situ Assessment

Sound in enclosures is described as a highly complex sound field, and its characteristics are affected by interdependent parameters, relating to the interaction between the excitation signals, sound sources, and the enclosure. In a general audio evaluation protocol, one aims for a systematic variation of certain physical parameters, that evoke a number of perceptual characteristics. When assessing real room acoustics, this approach is difficult; that is, systematic variation of single parameters of a real-field, e.g., varying the C_{50}, whilst the remaining characteristics are constant is impractical if not impossible. Therefore, most studies compare several real sound fields, e.g., different concert halls, and the perceptual observations are then contrasted to the physical characteristics of each concert hall, so that a correlative trend across all observations could be established. Such approaches formed the fundamental work in understanding architectural acoustics and the cognate perceived experiences [4, 7].

Unequivocally, the *in-situ* methods provide the most ecologically valid scenarios, as the human listeners experience the sound fields as they would occur in real life. Still, these methods encompass several limitations. It is somewhat impractical to instantaneously compare several acoustical fields in-situ, for example, the sound of a loudspeaker in several rooms. This imposes long test-to-test periods between the experimental stimuli, in this case the sound fields in each room, which is well known to provide erratic perceptual results, due to the limited auditory memory humans exhibit [48, 49]. Moreover, exposing assessors in a real environment introduces several non-acoustical factors in the evaluation, relating to the multi-modal experience of being in a space, aural room adaptation, as well as cognitive and personal cognitive factors of the listener, such as current mood, emotion, and taste of each listener [50].

Naturally, the *in-situ* evaluation was not followed in this work due to uncontrolled biases that may alter our observations when evaluating acoustics in typical, thus familiar environments, as well as the impractical challenges that would comprise for a comparative evaluation. Therefore, we focused in defining the appropriate acquisition, and presentation protocols so that a laboratory-based method could be followed. The process to define these is described below.

3.6 Laboratory Assessment

Laboratory based methods aim to directly overcome the limitations of *in-situ* assessment, by imposing higher control of the experimental factors and variables in a repeatable, robust, and blind protocol that allows simultaneous comparisons between audio material.

These methods comprise of two major parts: the *acquisition* of the field, where the experimental stimuli are captured, and the *presentation*, where the experimental stimuli are presented to the human assessors. The acquisition of sound fields, could be completed with computer *simulations*, where the signal, the source, and the enclosure are simulated based on their physical characteristics, and a computer model estimates the resultant sound field. *Measurements* could be also conducted in real fields, were the sound field created by a source in an enclosure is recorded at a specific location, e.g. the listening position.

The methods followed to acquire and present the experimental stimuli are equally important and drive the trade-off, between the ecological validity and the variable control in the experimental design. That is the extent to which the human responses evoked, within this framework, relate to the responses in the real-life aural experience, whilst imposing controls on the experimental factors.

These considerations highly influenced this work and the establishment of a perceptual assessment protocol to address our research questions. To further explore these challenges, a critical review of the published literature was conducted. This led to a series of pilot experiments in an attempt to define the most suitable acquisition and presentation schemes for evaluating sound fields within domestic rooms and automotive car cabins.

In the sections below, a brief description of this process is given, relating to the possible approaches that one could employ such as simulating and measuring sound fields. These include several observations through the reported experimental studies, that led to the requirements and proposals of an experimental methodology to be followed in this work.

Simulations

A number of algorithms exist that directly focus on simulating sound propagation in enclosures, i.e. modeling of the acoustical field. One could categorize these in *Geometrical Acoustics* (GA), *Wave Based*, and *Artificial* methods, as summarized in Fig. 8

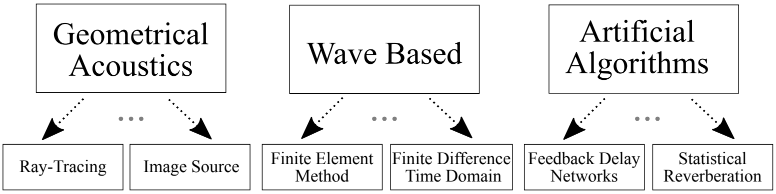


Fig. 8: Overview of acoustical models for sound propagation. Based on [51].

The first part of this work focused on simulating the acoustical fields of interest using GA and Wave-based models [47]. Acoustical models were developed based on two GA acoustics methods, i.e. image source and ray tracing techniques [51], using CATT-A. Moreover, as these methods are limited to low-frequency reproduction [47], relating to the size of the room model, a wave-based model, i.e. *Finite Difference Time Domain* (FDTD) method [52], was introduced for the low frequencies.

In these experimental studies, it was possible to simulate acoustical fields with systematic variable control that would be a major advantage for our study. However, the detail in which the physical properties of the enclosures could be defined in such models is restricted, typically dominated by the diffusion and absorption coefficients given in octave-wide frequency bands. In order to achieve a *plausible* [53] acoustical depiction of a real domestic environment, one needs to manually 'tune' the model in recursive experiential manner. Moreover, the model's source properties do not provide a real depiction of a loudspeaker but rather a crude simplification. Often the source is defined in terms of energy across octave bands and its directional characteristics are spatially-limited. The presentation of the field typically follows binaural auralization, i.e., presentation over headphones, which encompasses its own spectral and spatial limitations [10].

An acoustical field that this thesis addresses is the one that occurs in car cabins. Car cabins are physically an unconventional acoustical environment, exhibiting significant seat-to-seat variation and extreme modal behavior, which makes the physical quantification of the field highly challenging [34]. These properties of the field limit the possibility to adequately simulate cabin's acoustical behavior [54, 55] for perceptual assessment purposes. These approaches cannot easily provide adequate synthesis of an acoustical field of a car cabin over a wide-frequency range. That relates to the high computational demands of the methods, requiring oversimplifications of the model and violation of the methods' assumptions, such as the wave-length ratio and model's boundaries[3] in GA [51].

These challenges have potentially led to the observed lack of perceptual and ecological relevance of these approaches when simulating the sound fields of small enclosures, such as car cabins and ordinary rooms. It is understood that careful and systematic work, that compares real acoustical fields and computer simulations in parallel can improve these limitations. Yet, the efforts required did not fit our purposes and research goals. It was therefore decided that the investigation towards capturing these sound fields should be focused on the domain of measurements instead, as detailed below.

Measurements

Two basic techniques could be followed to capture the acoustical field using measurements, namely, the *Binaural* techniques and the *Spatial* techniques.

The *Binaural* techniques, aim to capture the pressure that would arrive to the ear of a human listener, including the effects imposed by the physical properties of the head and torso.

The *Spatial* techniques make use of microphone arrays, aiming to capture the spatial properties of the sound field, so that one can re-synthesize it at a later stage, using a spatial rendering scheme. There is a number of spatial methods that make use of multiple microphones, such as *Ambisonics*, *Spatial Decomposition Method*, and *DirAC*.

Each of these approaches encompasses its own advantages and disadvantages, following their implementation needs and their theoretical assumptions and basis. The limitations of the methods were identified in the literature and

---

[3] The considered wavelengths should be small compared to the mean-free-path in a room [11]

their applicability to our research interests was evaluated. The work focused on the two major measurement methods, which are described below in detail.

Binaural Techniques

Binaural techniques are well established in spatial audio research [56], due to the straightforward implementation and the solid theoretical background they operate on [58]. The term describes the methods relating to recording, synthesizing, and reproducing the signals that appear at the two ears of a human listener. In such approach one aims to capture the sound field that a human listener is exposed to in a way that the filtering effects of the head, pinna and torso are included, i.e., the *Head Related Transfer Function* (HRTF). When these signals are reproduced appropriately, the human auditory system appropriately decodes the imposed acoustic cues revealing a similar aural experience as being in the real field.

There are several ways in which binaural signals can be obtained, for example by *in-situ* measurements, where a recording apparatus is used to capture a real field, or by simulations where virtual environments are combined with anechoic HRTF resulting to virtual auditory display.

In this study we focused on binaural measurements. These measurements could be achieved by capturing the sound at the eardrum of a human listener, or by using artificial *dummy-heads*, or in some cases by using stereophonic microphone techniques, i.e., Jecklin Disk. Due to the practical challenges of capturing the field with human listeners, artificial *dummy-head* recordings are often employed. A *dummy-head* is exposed to the real field of interest and in-ear microphones are used to capture the signals at the two ears. These two signals are then synthesized and reproduced to the human listener via auralization techniques, e.g. headphone reproduction [56]. This process is shown in Fig. 9.

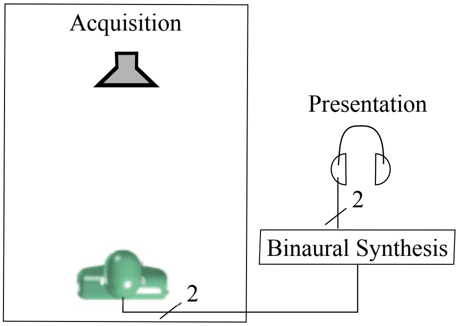


Fig. 9: Procedure for the acquisition and presentation of binaural signals.

In order to explore the applicability of these techniques to our research interests we first conducted dummy-head measurements in standard listening rooms and car cabins, as shown in Fig. 10. *Binaural Room Impulse Responses* (BRIR) were captured for multiple head orientations using *Binaural Scanning Methods* [57]. A number of acoustical conditions were obtained, for example, by altering the decay times of the enclosure, or exciting the enclosure using multiple sound sources. These were then reproduced binaurally over head-tracked headphones [58]. Informal listening tests compared the *in-situ* experience in these fields and the one delivered by the *Binaural Room Scanning* (BRS) system.

Fig. 10: In-situ measurements using a dummy head in ordinary rooms and car cabins.

In general, the results of the informal experiments indicated that binaural measurements and the subsequent synthesis adequately delivered the perceptual properties of the fields in terms of relative differences, e.g., between two loudspeaker systems. The methods could be used to identify the relative perceptual properties between certain acoustical scenarios. However, two major shortcomings were evident for our research purposes.

First, the spatial properties of the sound field were perceived as unfamiliar to the assessors, although they have been exposed to the real fields before. Moreover, during the binaural reproduction, spatial alterations that were apparent in the real field were perceived as timbral differences instead, i.e., coloration of the sound sources, as noted in previous studies of automotive audio evaluation [59] using BRS. The lack of externalization, i.e., *in-the-head* perception of the sources, resulted to alien and somewhat unnatural perceived experience of sound in an enclosure. The reported observations may relate to the use of non-individualized HRTF[4] synthesis and the fixed HRTF of the dummy-head. Summarizing the literature one could note that the extent to which individualized or compensated HRTF eliminate such effects is currently a field of research debate [53]. It was therefore unclear whether employing individualized HRTF would guarantee well externalized sound, rather than the identified distortions of the acoustical field.

Second, the reproduction over headphones provided inadequate low-frequency performance and lack of the whole-body-vibration [60] was also reported, as already identified in the literature [61]. This can be seen as an important limi-

---

[4] fitting or compensating for individual assessors' anthropometric properties

tation for a more natural aural experience, as it is known that low-frequencies are highly important in sound reproduction, especially in geometrically small enclosures such as ordinary listening rooms [60] and car cabins [33].

The findings of the experiments described above indicated a limited applicability of binaural methods to our research interests. The spatial nature of reverberation was highly anticipated in this work, yet, this property is not guaranteed to be preserved during binaural presentation. This may relate to a mismatch between the assessors' physical features and the dummy-head's, thus, a dissimilar HRTF to what the assessor would expect. Nevertheless, following these limitations, the spatial characteristics of the fields might be degraded, introducing technical artefacts and an unnatural experience, as reported in previous studies that employed binaural techniques in room acoustics assessment [10]. These observations steered our research directions towards methods that may overcome such limitations.

Spatial Techniques

Spatial methods aim to capture the acoustical field in the form of simultaneous measurements using multiple microphones, most commonly in three-dimensional orientations. These spatial measurements are used to analyze the acquired field, in a way that the field could be reconstructed in the laboratory by means of sound reproduction, e.g., over loudspeaker rendering[5]. These techniques overcome several limitations of binaural measurements. For example, the fixed HRTF on the captured field and the related limited externalization of the sound sources, as well as the limited low-frequency reproduction over headphones.

The published literature in the domain of spatial audio comprises of mainly two approaches, the physics-driven methods, and the perception-driven methods. The first type considers the methods that aim to recreate the physical sound field as accurately as possible. The second, relates to methods that focus on recreating the sound field as perceptually accurate as possible, i.e., delivering the same experience to the human listener as the captured field.

In the attempt to recreate the exact captured fields, most physically-driven techniques, e.g., (Higher Order) Ambisonics and Wave-Field Synthesis, introduce significant artefacts that are prominent to the human listener. These relate to non-accurate and perceptually colored sound fields and source properties [62], as well as a confined and impractical listening area [63]. Moreover, the reproduction frequency depends on the spacing between the reproduction loudspeakers, typically introducing *spatial aliasing* at high frequencies[6], which potentially introduce ill-conditioning [63] to listeners. These limitations affect the naturalness of the perceptual aural experiences, especially when the focus

---

[5] binaural presentation over headphones is also possible

[6] $>$ 4kHz for 5cm spacing between reproduction loudspeakers

is the evaluation of real fields at broad frequency bandwidths. The applicability of these methodologies to the project's interest has therefore seemed highly challenging and impractical.

The second family of spatial techniques that is eminent in spatial audio research, the perception-driven methods, aim to limit the influence of audible artefacts by following psychoacoustical principles. The most commonly used techniques in spatial audio are (1) *Directional Audio Coding* (DirAC) [64], (2) *Spatial Impulse Response Rendering* (SIRR) [65], and (3) *Spatial Decomposition Method* (SDM) [66]. These methods focus on recreating a more perceptually relevant experience, by preserving the perceptual qualities of measured sound fields in the best possible way. This is typically achieved by taking advantage of current psychoacoustical knowledge and the properties of the human auditory system. For example, the restricted spatial resolution of human hearing in complex sound fields and its enhanced sensitivity to timbral properties. These methods may however operate at the cost of the physical representation of the captured sound fields.

One of the most recent spatial acquisition and synthesis methods is the SDM [66]. SDM has been implemented following studies in concert halls that initially used DirAC [5] and later SIRR [6]. SDM employs an analysis of the *Direction Of Arrival* (DOA) in time segments, for each discrete captured sample of a spatial impulse response[7]. This information is combined with the pressure values from an omni-directional microphone, ideally located in the center of the array. The resultant vectors can be then used to spatially synthesize the field using a parametric rendering scheme. Recently, SDM was also used to perceptually evaluate studio control rooms [67], indicating its applicability for geometrically smaller spaces. SDM synthesis has been claimed to preserve the sound quality of the captured fields [68], whilst its implementation is straightforward allowing to presentation of the analyzed sound fields in any parametric rendering scheme [66].

Moreover, the method provides spatiotemporal analysis of the sound fields. This novel objective analysis of the captured sound field presents the energy distribution of reflections in terms of time and direction [69]. The literature has primarily focused on the temporal measurable quantities of reverberation, due to the limited capacity to physically characterize the spatial properties of the field. Thus, such metric could be highly valuable when investigating the perceived spatial aspects of the fields. The basic principles of SDM are shown in Fig. 11.

As SDM was first developed for concert halls, we first sought the applicability of the method in car cabins, the most diverse acoustical field of this project. A number of acoustical alterations[8] of a car's interior were conducted and the

---

[7] Impulse Responses captured simultaneously with a multi-microphone array, such as 3D microphone probe

[8] see Paper [C]

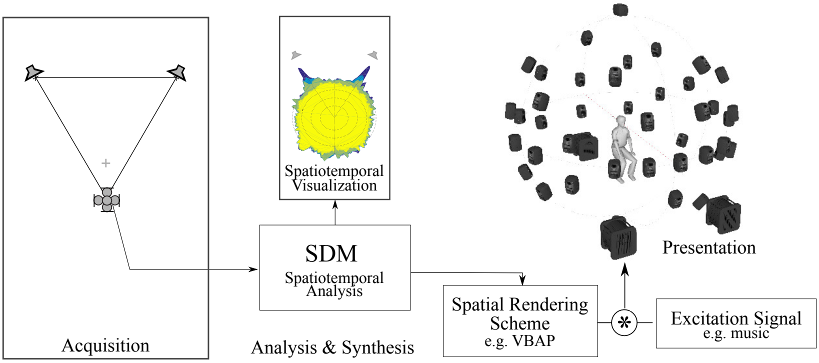


Fig. 11: Basic overview of Spatial Decomposition Method.

resultant fields were captured and analyzed using SDM (see Fig. 12). These studies led to a proposed adaptation of the method to allow for the analysis of near-field sources. The process followed to acquire the acoustical fields in the cabin (see Fig. 12), including the results and proposed adaptation of the method were summarized in Tervo et al. [70] (see Paper [E])[9].

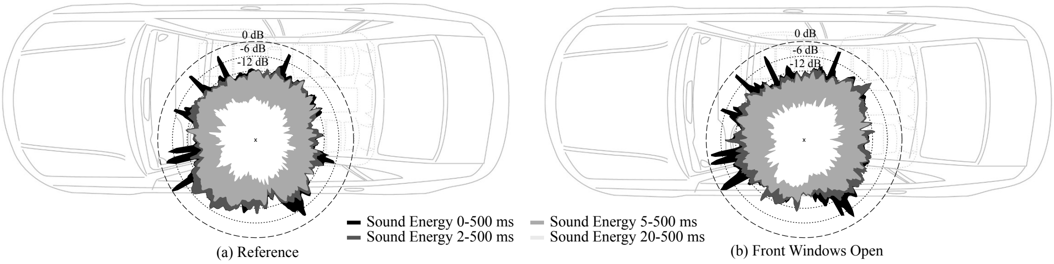
In order to verify the performance of the method, several acoustical conditions of a car cabin were analyzed using SDM and their deviation from the reference condition was reviewed. Figure 13 shows the energy distribution over time in a car cabin, equipped with 17 individual transducers. On the left (a) the figure presents the analysis for the typical production car with no modifications, the reference condition, and on the right (b) the figure presents the energy distribution in the cabin when the front windows of the vehicle were opened. The accuracy of the analysis cannot be physically verified as a benchmark of these methods is not straightforward. Still, the relative differences between the reference and an acoustical condition can be contrasted. As it can be seen in Fig. 13, the direct sound (0-500 ms) is identical between the two conditions, whilst the lateral energy that could be related to the reflections of the windows, is significantly lower when windows are open compared to the reference, as one would expect. This signifies that the SDM is able to analyze such fields and yield their spatiotemporal behavior of the field, which seems to follow the theoretical explanations.

A series of informal listening experiments was also conducted where a number of acoustical conditions in car cabins were perceptually evaluated. The method was found to provide an externalized sound field, as it made use of a loudspeaker array, and it was capable of acquiring and adequately reproducing the frequency range of interest (20-20k Hz). These advantages directly addressed the challenges raised with previously used methods. The acoustical

---

[9] where the the author of this thesis was a contributor

Fig. 12: Example of measurements that were conducted in car cabins, where the interior was physically altered.



(a) Reference

(b) Front Windows Open

Fig. 13: Spatiotemporal analysis for two conditions in the car cabin, indicating the ability of the method to identify the spatial distribution of energy over time in car cabins. Adapted from Paper [B].

conditions presented evoked perceived sensations in both spatial and timbral domain, as reported in Tervo et al. [70], and in more detail in [34], Paper [B]. It was also apparent that no technical inadequacies were audible, e.g., artefacts resulting to perceived noise or spatial unnaturalness, which is commonly observed when spatially reproducing a sound field [10].

3.7 Evaluation - Decoding Human Perception

The last part of a perceptual assessment framework includes the evaluation protocol where the stimuli are assessed and characterized by human assessors. The evaluation process typically follows the form of a listening experiment were the assessors are asked to quantify their perceived experience. These observations are then contrasted to measurable physical features of the stimuli, aiming to identify the relationships between them.

As it has already been discussed, reverberation is a highly complex sound stimulus, described by a number of interdependent physical and perceptual features. Thus, it is rarely possible to directly relate a physical measurable quantity to the perceived experience in such investigations, especially when real acoustical fields are in question. The perceived sensations in such experimental designs would typically include a wealth of both timbral and spatial aspects, relating to the enclosure, the source, and signals. This adds another layer complexity to the experimental analysis, decreasing the ability to relate the perceived attributes (i.e. perceptual domain) to acoustical features of the field (i.e. physical domain) and the global experience (i.e. affective domain).

In order to decompose the multidimensional character of the perceived sound in enclosures several studies employed qualitative methodologies [4], e.g., structured questionnaires and interviews, where listeners described their perceived experiences on paper, or verbally. These studies [10] have been fundamental in sound quality research and they further demonstrated the challenges of defining the perceived sensations in single and quantifiable descriptors that one can relate to physical and pragmatic quantities. These observations were presented in Paper [A], highlighting that the vocabularies and attributes used in the published literature include ambivalent definitions and labels, making the generalization of the results a complicated matter.

It is agreed that employing human perception as a measuring instrument imposes a number of biases that the experimenter needs to overcome. This may include own references and judgments, as well as cognitive influences, e.g., mood. These potential biases could be limited by employing controlled and repeatable evaluation protocols. That is, a protocol that follows the contextual factors and applications of interest, resulting in domain-specific approaches.

Several audio evaluation protocols have been proposed over the last decades, that one can categorize in two groups: (1) the *Attribute*-based, and (2) the *Basic Quality* methods. In attribute-based methods, a list[10] of perceptual attributes is used to quantify the perceived sensations, what is referred to as *Perceptual domain* by FM. In the contrary, Basic Audio Quality methods [37] seek to quantify the perceived experience in a single measurable quantity, relating to personal preference and likeness, what FM describes as an *affective* response. Based on these principles several evaluation protocols have been established and standardized including recommendations for the perceptual assessment of small degradation of signals [37], coding schemes [38], and sound reproduction in ordinary rooms [42]. There is currently no standardized procedure for perceptually evaluating room acoustics and their properties.

In an attempt to follow a well-structured and repeatable evaluation framework, recent studies [5, 6, 71] have adopted *Sensory Analysis* (SA) methodologies [72], commonly found in food and wine industries. The major advantage of these techniques lies on their ability to mathematically relate the physical, perceptual, and affective domains in the same experimental design and in consequence in a unified statistical analysis. That is, a multivariate decomposition of the stimulus' properties in all three domains, i.e., physical, perceptual, and affective, in each domain separately, as well as merging the domains in a single analysis. It is evident in the literature that these techniques allowed a novel characterization of complex and multidimensional entities, whilst avoiding com-

---

[10] pre-selected or elicited

monly encountered challenges, such as the erratic interpretation of verbal descriptors, and uncontrolled personal sentiments and judgments.

The main research scope of this thesis was to identify and quantify the perceptual aspects that dominate the properties of sound in the rooms of interest and relate these, when possible, to physical alterations of the field, i.e., measurable quantities. As the perceptual aspects are multiple and not known *a-priori*, the evaluation protocol to be employed, should be capable of decomposing the perceived experience in separate attributes. Still, the protocol should provide means to investigate the relationships between the identified attributes and the physical quantities. These requirements are met by several verbal elicitation SA methods, e.g., *Quantitative Descriptive Analysis* (QDA) [72]. However, these are laborious procedures, requiring multiple experimental sessions. Thus, an adaptation of these methods was considered in this study, the so-called rapid SA methods [73, 74]. The method chosen, the *Flash Profile* (FP) [75], was found to adequately meet the requirements of this exploratory investigation, due to its limited experimental time required, the free verbal elicitation process, and the absence of training procedures or requiring reference stimulus, i.e., an optimal sound field. The details and rationale are given in Papers [B] & [C].

Perception, Expectations, & Multimodal Processing

Every enclosed environment has a sonic identity [43]. Humans are evidently able to identify the characteristics of their surrounding environment based on information contained in the reverberant field [76–78] alone – in the absense of visual input. However, in realistic environments this process involves both visual and auditory information processes, where the visual representation accounts for most [79]. It has been shown that visual information distorted human judgment on auditory tasks including estimating auditory distance [80], spatial impression [81, 82], and spatial width [83, 84]. Although humans are not often aware of these processes, especially when sound matches expectation [26], it is well known that in a mismatch between aural and visual cues, one would dominate, in an attempt to establish a known mental scenario [85], for example based on memories of architecturally similar spaces [83, 84].

In fact, the human auditory system processes the information contained in a reverbrant field to primarily serve a fundamental and evolutionary purpose, what is known as *Space Perception* [85]. That is, the ability of humans to identify and define the properties of their surrounding environment [86], in order to establish safety [85] and adapt their behavior [77, 87] to the environmental factors. This involves multiple low-level processes and mechanisms, from several modalities, all of which contribute to a mental representation and spatial map of the surrounding environment.

It is therefore critical to respect that once a listener is exposed to a large concert hall, a car, or a living room his/hers perceptual and cognitive auditory processes might be modified due to memory and expectations [50, 83], even before auditory information is perceived. In consequence, the human assessor's judgments and sentiments may influence his/her responses to the acoustic properties of a signal or an experience.

In an experimental setting, e.g., in perceptual studies conducted using virtual environments, this may require the assessment of an acoustical field to be accompanied by a realistic visual and spatial experience, or a way to neutralize the visual input when dissimilar environments are to be evaluated. Controlling the visual inputs, e.g., by blindfolding assessors, may not reveal a true representation of a realistic experience, yet, it controls the interaction between visual and aural inputs, whilst space perception is aided by aural information.

In this work, several steps were taken to address these factors. This included a controlled experiment, where the experimental apparatus, room, purpose of the experiment was kept unknown to the subjects. The listening experiment was conducted in dark conditions and the assessor was separated from the apparatus with an acoustic curtain, both during and in-between sessions. It is understood that these aspects of reverberation and its properties are fundamental in perceptual evaluation and they must be considered in an experimental design where an auditory response is to be evoked.

4 General Remarks

4.1 Limitations & Considerations

In the previous sections the major physical and perceptual aspects of reverberation were presented. It has been noted that several properties of the field, e.g., the decay times, the strength of reflections, and the relative energy between early and late, could be quantified mathematically. These parameters seem to alter the perceived experience, relating to multiple sensations in a non predictable manner. This interdependence is also mirrored in the perceptual domain, as the perceptual attributes are clearly not independent to each other. These findings indicate that it is currently impossible to create an accurate perceptual profile of an acoustical environment based on physical metrics alone, or *visa versa*. This disconnection between the physical and perceptual domains limits the ability to determine the relations to the affective properties of an acoustical environment, that is, the attributes and parameters that could characterize the overall human experience in a given environment. Following these observations, the latest studies in the domain follow the conjoint investigation of both the physical and perceptual properties of acoustical fields.

It is has also been apparent that the majority of the studies on reverberation focused on geometrically large environments, typically described by prominent prolonged decays and the dominant sense of *reverberance*. Our research scope seeks to study these effects in smaller acoustical enclosures that one encounters every day, i.e., ordinary residential environments and automotive car cabins. In the lack of perceptual studies in our domain of focus, it was a natural choice to initially investigate the findings within spatial audio, architectural acoustics, and psychoacoustics, and seek their applicability to our research interests. A limited number of studies exist that directly assessed the properties of small rooms as a complete sound field, in contrast to the plethora of research on large performance spaces. As previously discussed, these spaces encompass acoustically divergent fields compared to typical listening environments that may introduce several limitations in translating their results between the two domains. To explore these assumptions, a literature review was conducted (Sec. 2-3 & Paper [A]) and attempted to identify common paths of reverberation research. It further assessed the extent to which findings in large performance halls translate to geometrically and acoustically smaller environments and stipulated the factors that may drive their relationships.

The primary findings highlighted the physical differences between the two domains that may influence the perceived sound experience. Based on this evidence, it seemed unlikely that the previous findings in spatial audio research and performance halls studies could be translated to our research interests. Three major factors were found drive this divergence, namely:

1. Source & Signal Properties
2. Enclosure Properties
3. Human Biases - Expectations & Multimodal Processing

Moreover, several considerations of the state-of-the-art in perceptual assessment of room acoustics were stipulated, and the challenges imposed by these factors were described above in detail. In a series of exploratory studies, the advantages and shortcomings of the major approaches were identified. These motivated the design of a new experimental protocol, which aimed to address the major shortcomings identified in the literature. The proposed protocol comprised of a presentation of the sound field in a perceptually relevant way, whilst the evaluation procedures were controlled, repeatable, and blind, based on a robust scientific framework.

5 Thesis Overview

The main topic of the work presented in this thesis is the perception of sound in enclosures and the effects of reverberation[11] *per se*. The research focus of the current thesis was in geometrically small listening environments that one encounters in every day life, such as domestic rooms and car cabins.

In order to address the specific research interests in such a wide subject as reverberation, a deductive, top-down approach was followed. The effects of room acoustics properties and human perception were first investigated and directed the research focus on acoustical fields that could address our research questions, as graphically shown in Fig. 14.

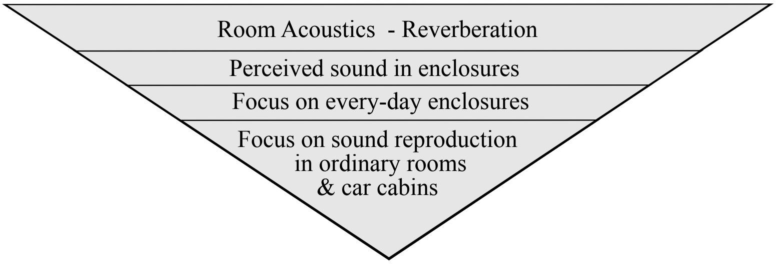


Fig. 14: Rationale and deductive top-down approach of this thesis.

Part I of this thesis presented the rationale and background of this work, leading to the identified limitations, and the experimental methods that were followed in these investigations. In the next part of the thesis, the contributions are presented in four articles, and supplementary material is given in the appendices. In the following sections these contributions are briefly presented, and the major conclusions and limitations of the studies are discussed, leading to proposals for future work.

5.1 Contributions & Dissemination

The main part of this thesis, Part II, comprises of a collection of four papers. The first two papers focus on the literature pertaining the perceptual evaluation of room acoustics, and the design of a novel experimental methodology that met the requirements of this work and could address the proposed research questions. The last two papers, include the results of implementing the methodology in automotive audio systems, and in ordinary residential listening rooms. Paper [E] and [F] contain related work to this thesis, given as supplementary material in the appendix. The contributions are summarized in Fig. 15.

---

[11] In this work, we refer to reverberation as the result of the interaction between a sound source and an enclosure.

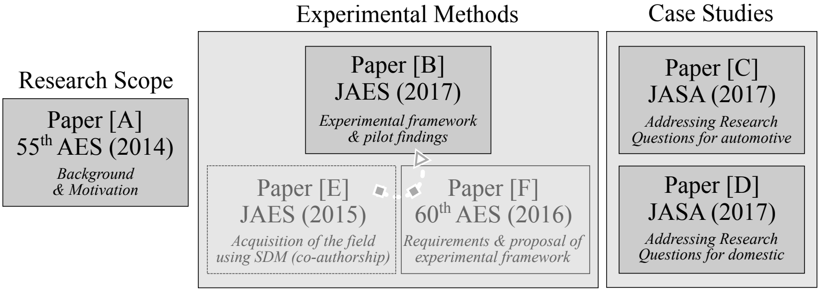


Fig. 15: Overview of the work included in this thesis, in a chronological continuum.

In order to address the research questions proposed in Sec. 1.1, we employed an iterative active research process [88]. Concurrent cycles were followed to plan, act, observe, and reflect on findings so that the major research questions were investigated in the best possible way. These cycles focused on understanding the current state-of-the-art, which set the *foundations* of this investigation. This allowed us to identify the limitations noted in the literature and design a novel *experimental method* that could overcome these and address our research interests adequately. Two *case studies* followed, which investigated the perceived sound fields in car cabins and domestic ordinary rooms, respectively.

The initial work that was described in the previous sections, has been detailed in Paper [A]. Motivated by these findings, Paper [B], detailed the requirements for evaluating sound in the enclosures of our research interest and presented a new experimental framework. Having established a perceptual assessment framework, two case studies followed, investigating the perceived effects within car cabins and ordinary listening rooms. In Paper [C], we implemented these experimental procedures, to identify the effects of the main acoustical compartments inside a car cabin on the perceived sound field. This led to Paper [D], a study that focused on investigating the perceptual aspects of reproduced sound in the most used listening environments [14], the residential ordinary rooms.

The research undertaken in this PhD project is presented below, depicting the directions followed during this work, ordered in a chronological continuum. The summary of the contributions and their linkage to the research questions is summarized in Fig. 16.

1. **Paper A - Perception of Reverberation in Small Rooms: A Literature Study**

   This work presented a comprehensive literature review in several domains of perceptual audio, relating to room acoustics and reverberation *per se*. The main aim of this study was the identification of trends and similarities across the published literature, in disciplines relating to room

	PLAN	ACT	OBSERVE	REFLECT	DELIVERABLE
FOUNDATION	Can published literature answer our RQ1-3?	Identify & assess literature and current practices	·Mainly concert hall studies ·Reverberation is multidimensional ·Similar trends across studies ·Assessment protocols in-situ or Laboratory-based	·Physical & Perceptual characteristics unlikely to be similar to small rooms ·Physical-Perceptual relations vary ·Current perceptual assessment focus on full complete sound field evaluations	PAPER A Background & Applicability to RQ
EXPERIMENTAL METHODS	How can we evaluate sound in small enclosures?	Identify current practices & seek for applicability	·In situ assessment + Questionnaires ·In Laboratory + Comparisons	·Small rooms are difficult to simulate ·In situ is impractical & includes biases ·Binaural lacks of spatial attributes ·Studies elicit technical attributes, relating to the method (e.g. slow head-tracking)	PAPER B Design of Assessement Protocol
	What are the requirements for perceptual assessment of geometrically smaller enclosures?	·Identify major factors that differ for these enclosures ·Conduct experiments to evaluate possible methods	*Experimental Requirements:* ·Multivariate protocols in design and analysis ·Source & signals - constant ·Enclosure's properties - variable ·Preserve spatial & temporal features ·Control visual and memory biases	·Simulations not perceptually adequate ·Binaural techniques lack externalization & low frequency performance ·Multimicrophone techniques include audible artefacts and high demands ·Evaluation protocol must allow simultaneous comparisons & multivariate analysis	
CASE STUDIES	Research Q1 Research Q2 Research Q3	·Systematically vary factors that are likely to affect the sound field ·Seek for relationships between physical and perceptual aspects	·Strong and multiple reflections ·Dry environment but dinstict field ·High number of sources ·Cannot conduct between car studies with sources being a constant ·Strong modal behavior	·Evaluation of one cabin (with-in) ·Alteration of major compartments of car's interior - 12 conditions ·Initial verifications for the assessment protocol included in design ·Physical/objective metrics limited	PAPER C Investigation of sound in car cabins
			·Similar early reflections ·Sources can be controlled ·Variation of room's properties possible ·Comparison between rooms possible ·Modal low frequency behavior varies significantly between rooms	·Systematic Variation of decay time in one room (with-in analysis) ·Comparison of four residential rooms (between analysis) ·Objective acoustical metrics available	PAPER D Investigation of sound in ordinary rooms

Fig. 16: Summary of the contributions, based on the active research cycle [88].

acoustics. That is, evaluating the possibilities that knowledge from other domains could support our research scope, to perceptually characterizing small room acoustics. As studies directly addressing the volumetrically smaller environments, e.g., car cabins and ordinary rooms, were limited, the work evaluated the scientific literature in spatial audio reproduction, psychoacoustics, acoustical measurements and the plethora of studies linking concert hall acoustics to perceptual aspects of the sound field. This provided a holistic overview of the most common perceptual attributes relating to acoustical properties of the fields (RQ1). Several studies identified relationships between physical parameters of the sound fields and the major perceptual attributes that they evoke (RQ2), as well as linkage and inter-relations of perceived sensations (RQ3). Based on these findings, it was suggested that smaller environments comprise of physical limitations that would affect perception in a dissimilar way than in the domain of spatial audio and concert halls acoustics. Thus, the applicability of the current literature to smaller rooms seem challenging. The study summarized the identified requirements to address our research questions, and several proposals for a novel experimental design were made.

This work was presented at the 55th International Conference of Audio Engineering Society, on Spatial Audio, in Helsinki, Finland, August 2014.

2. **Paper B - A Rapid Sensory Analysis Method for Perceptual Assessment of Automotive Audio**

   Following the findings and suggestions of Paper [A] and a series of experimental studies, this work extended the initial proposals of a novel assessment framework for evaluation of sound in the acoustical fields of interest[12]. The paper presented the reviewed the current perceptual evaluation protocols in automotive domain and stipulated the limitations of the state-of-the-art. A research framework was then proposed to directly address our research questions. The method made use of SDM for acquisition and presentation of the sound fields in the laboratory, aiming to overcome issues raised in the literature, and our exploratory studies.

   A rapid sensory analysis method, the *Flash Profile* (FP), was followed for the perceptual characterization of the sound stimuli, where individual vocabulary was used to identify and quantify the perceived sensations. The experimental framework was then implemented to perceptually assess automotive audio. The initial results were discussed, depicting the advantages and limitations of the proposed methods.

   This work has been accepted for publication in the Journal of Audio Engineering Society, Vol 134(1/2), 2017. Parts of this work can be found

---

[12] first described in (Paper [F])

in Paper [F], describing the work presented in Proc. of the 60th International Conference of Audio Engineering Society, on D.R.E.A.M.S, in Leuven, Belgium, February, 2016.

3. **Paper C - Perceptual Aspects of Reproduced Sound in Car Cabins**

   In this work, the perceptual characterization of the acoustical fields in car cabins was addressed, by implementing the proposed methodology, described in Paper [B].

   Automotive car cabins encompass a unique and somewhat hostile acoustical environment, comprising of challenges in both its physical and its perceptual characterization. This is in fact, one of the major challenges in achieving high quality sound reproduction in cars. The main purpose of the study was the investigation of the extent to which acoustical alterations in the car cabin influence the perceived sound field.

   As comparing different car cabins whilst the reproduction system is identical was not possible, one car cabin was acoustically modified and twelve acoustical fields were used in a listening experiment. Their perceptual effects were quantified by expert assessors using *Flash Profile.* The perceptual attributes relating to the acoustical modifications were identified and analyzed, which highlighted the influence of several compartments on the perceived experience, even in such a temporally-limited and dense reflective environment. This work presented the most salient perceptual attributes (RQ1) that related to acoustical fields within a car's interior. The quantification of these attributes in a ranking protocol indicated the perceptual importance of certain physical alterations (RQ2), and possible relationships (RQ3) between the two were proposed.

   This work has been accepted for publication in the Journal of Acoustical Society of America, to be published in 2017. Parts of this work were presented at the 171th meeting of the Acoustical Society of America, Salt Lake City, May, 2016

4. **Paper D - On the perception and preference of reproduced sound in ordinary listening rooms**

   In this work, the experimental framework described in Paper [B], and used in Paper [C], was modified for evaluation of ordinary listening rooms. The practical and statistical limitations identified in the previous works were addressed allowing the interpretation of perceptual, hedonic, and physical characteristics of these sound fields. A series of in-situ measurements was conducted, leading to a database of nine acoustical fields, from four standard listening rooms [42]. This included the response of four listening rooms, as well as a set of six sound fields originating from one room that was physically modified. Ten expert assessors characterized and quantified their perceived sensations to these fields, using Flash Profile. These were then contrasted with measurable physical features of the fields and the assessors' preference. A multidimensional analysis was performed and depicted the most prominent perceptual attributes evoked by the presented fields (RQ1) and contrasted these to physical characteristics that objectively described these sound fields (RQ2), and their relations (RQ3). The preferences of these sound fields were also collected and discussed.

The results highlighted the remarkable importance of reverberation in small, ordinary, residential listening rooms, on the perceived listening experience. The current findings depict a limited perceptual space for these sound fields, where the decay time dominates the assessors' evoked sensations.

This work has been submitted to the Journal of Acoustical Society of America, in November 2016, and as of December 29, 2016, it is currently under peer-review. Parts of this work were presented at the 171st Meeting of the Acoustical Society of America, Salt Lake City, May, 2016.

The following works are provided as supplementary material for completeness.

5. **Paper E - Spatial Analysis and Synthesis of Car Audio System and Car Cabin Acoustics with a Compact Microphone Array**[13]

   In this exploratory work, we investigated applicability of acquisition and presentation of the automotive sound fields using Spatial Decomposition Method. A series of experiments is reported, and modifications of the initial SDM algorithm were proposed to operate for near-field sources in highly reflective environments. The reported experiments include physical and perceptual studies, confirming the applicability of the method for assessing car cabin acoustics and automotive audio. Moreover, using SDM analysis it was possible to physically characterize the properties of the fields in a novel representation that conjointly presents spatial and temporal properties of the field. Such approach was found to be highly anticipated in automotive audio. These initial investigations led to the design of a purpose-made laboratory and experimental apparatus which was used in Paper [C] to perceptually evaluate automotive audio over loudspeakers.

---

[13] The author in this paper is only a partial contributor.

6. **Paper F - A method for Perceptual Assessment of Automotive Audio Systems and Cabin Acoustics**

   In this paper the initial experimental framework for assessing automotive audio using SDM and rapid sensory analysis was presented. This work combined our findings in Paper [A] and Paper [E] in which we discussed the requirements for assessing these environments and we reported the applicability of SDM in the acquisition and presentation of the sound fields. In this work we introduced rapid sensory analysis, and defined the requirements for the evaluation protocol. These were met by Flash Profile protocol, which is described in detail. The practical implementation of the method was then presented and pilot results were shown. This paper was later expanded and accepted for publication in the Journal of Audio Engineering Society (2017) (Paper [B]).

6 Summary & Conclusions

In this thesis, we investigated the influence of acoustical properties of typical listening environments on human perception. These investigations addressed three major questions relating to (a) the perceptual attributes that could characterize these fields, (b) the physical properties of the fields that are likely to influence human perception, and (c) the relationships between the physical and perceptual properties of the fields. In order to address these, the thesis also dealt with the appropriate methods to perceptually evaluate room acoustics and a new experimental framework was proposed and implemented.

The following sections summarize the main findings of this thesis. The first part, Sec. 6.1, discusses the research foundations and presents the related literature, leading to the rationale and motivation of the study. The requirements and considerations that drove the development of a perceptual assessment methodology are discussed in Sec. 6.2, and the key findings of the two case studies are presented in Sec. 6.3. In the last part, the possible limitations of this work are described and future directions are stipulated.

6.1 Research Foundation & Motivation

The first contribution, Paper [A], reviewed the literature pertaining the perceived sound in enclosures, and assessed their applicability to our research interests. The literature showcased a wealth of perceptual attributes for describing the aural experience in enclosures (RQ1), mostly relating to performance spaces. Yet, most studies provided ambivalent definitions and labels for these sensations, which degraded our ability to summarize the findings across studies based on semantic equivalence. It was apparent that several research streams used identical attribute labels to describe different perceptual sensations, and *vice versa*. In Paper [A], the studies were summarized in terms of given definitions. This allowed us to identify the commonly evoked perceived sensations and summarize their interrelations as well as the proposed physical parallels.

Nevertheless, the study highlighted the need for a common vocabulary to perceptually quantify aural experiences. This finding is not novel in audio evaluation. A number of vocabularies have been proposed recently, comprising of well defined perceptual attributes that one can use for sound reproduction [89], spatial audio properties [90], and virtual environments [91]. There is currently no vocabulary available relating to the evaluation of the perceived sound in enclosures, or reverberation *per se*.

In this thesis, we aimed to identify the most salient attributes that could describe the perceived sensations when experiencing the acoustics of domestic and automotive environments. As studies in these spaces were limited, the initial review focused on the published literature in perceptual evaluation of performance spaces and the cognate spatial audio research. This review has shown that these diverse acoustical fields seemed likely to evoke sensations in a dissimilar manner, than in smaller enclosures, that could relate uniquely to the physical properties of each field (RQ2–3). It was therefore not clear whether these attributes are applicable to smaller spaces, i.e., to our research interests.

The limitations in translating the accumulated knowledge to geometrically smaller environments, motivated the exploration for a novel experimental design in which the relevant attributes could be identified and quantified perceptually. The state-of-the-art in perceptual assessment of sound in enclosures was then examined and the influencing factors were identified, as reported in Sec. 3 and further in Paper [A] and [B]. These studies highlighted a fundamental *trade-off* in perceptual evaluation protocols. That is, the balance between the ability of a method to evaluate as natural an aural experience as possible, against its capacity to impose high variable control in the experimental factors.

6.2 Perceptual Assessment Framework

As reverberation is a multifaceted and complex entity it is difficult to control its features, i.e., the experimental variables, independently. Thus, the majority of studies focused on *in-situ* evaluation of real environments. There, the commonalities between several environments were drawn in both the perceptual and physical terms and their interpretation was attempted in a comparative manner. It was clear, however, that whilst maintaining the *ecological* validity and naturalness of the sound field, several non-auditory factors were introduced. These include visual and spatial expectations, as well as affective judgments relating to individual's taste and memory. Similarly, the major contextual factors, e.g., the sound source, the signals, and the enclosure's differences, were not always possible to be systematically controlled, limiting the ability to employ repeatable and scientifically robust evaluation procedures.

Laboratory-based methods address these challenges by recreating the aural experience in a neutral experimental space, where the human assessors could simultaneously assess the stimuli presented. These experimental methods allow for repeatable protocols including systematic controls on the experimental variables and factors. Hence, they effectively limit the influence of uncontrolled biases that may occur, relating to human perception and cognition. This apparatus, however, may limit the generalization of the results. That is, the results and observations of laboratory studies could be challenged as the perceived experience in the laboratory may not accurately reconstruct the *in-situ* experience.

In this work, we followed an *active research* [88] approach in order to define the requirements of the perceptual evaluation protocol. This allowed us to identify a method that could best address our research questions by targeting the limitations of the current methodologies. During this process our aim was to follow a protocol where evaluation procedures were controlled, repeatable, and blind, based on a robust scientific framework. Moreover, it was our aim to present the sound fields in the most perceptually relevant way, so that both the spatial and temporal characteristics of the fields were preserved to be evaluated by human assessors.

That led to a proposal of an alternative perceptual assessment protocol that included the acquisition, presentation, and evaluation of sound fields; detailed in the second contribution, Paper [B]. The method made use of a hybrid approach where real sound fields were captured *in-situ* using measurements, and spatially reproduced in a neutral environment over a loudspeaker array, shown in Fig. 17. A comparative and blind perceptual evaluation protocol could be therefore followed, where the major influencing factors, i.e., the source, the signal, the enclosure properties, and possible cognitive biases, could be experimentally controlled.

To address the multidimensional character of the sound stimuli during the listening experiment, a rapid SA protocol, the *Flash Profile* was adapted for audio evaluation and used to identify and quantify the perceived sensations of human assessors. This protocol allowed for the perceptual assessment of several acoustical settings, between and with-in enclosures, aiming to decode the complex character of the physical features within the fields and the evoked perceptual sensations.

The experimental framework that was followed in this work seems to overcome previously noted challenges in perceptual evaluation of room acoustics; related to acquisition, presentation and evaluation of such fields (Sec. 3). These include the non-externalized sound when presenting the sound fields over headphones, as well as the audibility of technical inadequacies and artefacts that limit the ability of the human assessor to relate the experience in a real life scenario. Here, no attributes relating to technical issues were elicited when

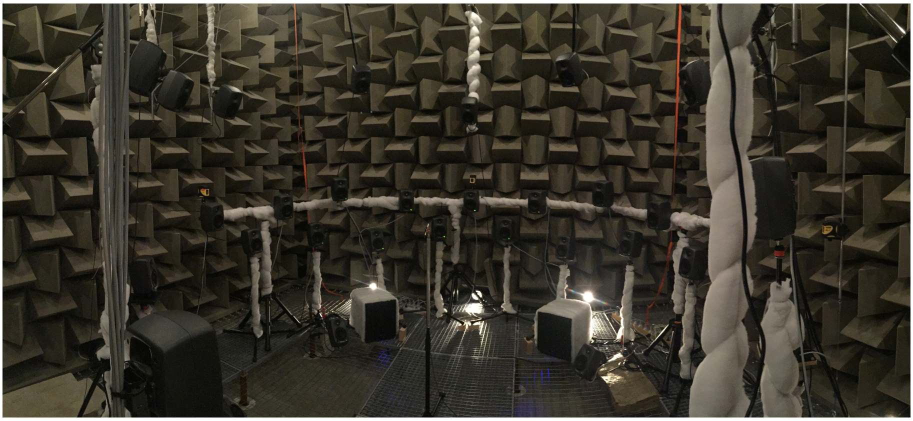


Fig. 17: An example of the experimental apparatus installed in the anechoic chamber. This apparatus was used in the pilot investigations for automotive audio.

using the proposed framework, in contrast to previous laboratory studies [10]. This is a remarkable result considering that all assessors were product-experts, i.e., critical listeners, and the scales and attributes were not fixed or defined.

Flash Profile has been shown to include several advantages that support the evaluation of acoustical sound fields. The method requires no reference, which is highly useful when comparing both with-in and between stimuli in a single experimental setup, or when novel and unique systems are in question. In addition, its rapidity meets the demands of the automotive industry without affecting the underlying statistical analysis. It was shown that the physical, perceptual, and affective responses can be merged in a summative statistical representation, mirroring the multidimensionality of these sound fields. Furthermore, FP makes use of individual verbal elicitation procedures, where each assessor defines his/her perceived sensations using own descriptors that are then used to quantify the magnitude of the perceived experience in a uni-dimensional manner. This effectively eliminates the influence of linguistic and ambivalent interpretations of the perceptual attributes, which were often noted in similar studies.

All in all, this work highlighted the possibilities of the Sensory Analysis methods for perceptual assessment of room acoustics, as well as the ability of SDM to preserve the acoustical properties of the acquired sound fields. Having new analysis tools, such as the spatiotemporal analysis provided by SDM, could potentially lead to new investigations focusing on the spatial distribution of reflections in a systematic way, and better understanding of the spatial aspects of these fields.

6.3 Case Studies

The third and fourth contributions, Paper [C] & [D], implemented the proposed experimental framework in order address the proposed research questions (Sec. 1.1).

Paper [C] investigated the perceived effects within car cabin acoustics. Although the car cabin has lately become a major listening space that humans enjoy music in [14], the physical and perceptual characteristics of this environment are not well understood, and the research in the domain is very limited. The complex geometrical and acoustical properties of a cabin form a unique sound field. These acoustical features result to unreliable physical description of the field, requiring automotive audio manufacturers to rely heavily on perceptually-driven methods. That is a process where manual sound tuning by an audio engineer is followed, in order to design, optimize, and deliver a pleasing experience to the listeners inside the cabin. Understanding the relationship between physical and perceptual aspects of these fields will be a significant advantage in this process.

This study has therefore focused on the effects of the acoustical properties of the cabin, by imposing physical alterations on the interior of a car cabin. This with-in cabin evaluation allowed us to investigate the effects of specific alterations of the field on the perceived experience. The results indicate that even slight alterations in the cabin, e.g., by reducing the reflective properties of a boundary, have a noticeable effect in the perceived experience. It was also shown that changes of the decay times, e.g., by altering the absorptive properties of the cabin, were observed by human assessors even at these marginal values (i.e. 80 ms). Such findings highlight the divergence of these environments compared to typical rooms and the incapacity of temporal based metrics to describe such fields physically. The identified attributes (RQ1) related to timbral alterations, such as the perceived *Bass* and *Brightness*. Spatial attributes were also evoked related to the sense of *Ambience*, *Width & Envelopment*, and *Image Focus*. Several relations (RQ3) between the physical (RQ2) and the perceptual properties of the field were identified and summarized in the article. For example, it was shown that the decay times influenced the extent to which the field was perceived as more ambient, whilst the windscreen properties had a remarkable effect on spatial and timbral aspects of the perceived sound.

The last contribution, Paper [D], investigated the effects of the acoustics in ordinary residential rooms. Aiming to evoke several perceptual attributes that are likely to occur in such environments, nine acoustical fields were captured in four standardized [42] ordinary listening rooms. Previous studies focused on the effects of the interaction between the loudspeaker's properties and the room, as well as the audibility and influences of the first reflections. Here, these factors were kept constant. The signals, sources, and the first reflection points were not modified[14], aiming to focus on alterations of the field that relate to the acoustical properties of the enclosure.

Using Flash Profile, a comparative evaluation was conducted where the sound fields were evaluated. These fields were captured in four standardized ordinary listening rooms, including six fields captured in a single room whilst its physical properties were modified. In this study, it was possible to combine both the physical, perceptual, and affective domains (discussed in Sec. 2.2) in the same assessment protocol, using a rapid sensory analysis evaluation and multivariate statistical analyses. These processes are described in detail in the manuscript, Paper [D].

Two major dimensions were found to explain the perceived differences of the evaluated rooms. The dominating sensations (RQ1) related to the perceived *Reverberance*, *Width & Envelopment*, sources' *Proximity*, as well as the perceived low frequency content, i.e., *Bass*. The results indicate that alterations of RT_{60} have an important effect on the perceived sensations (RQ2) as well as the assessor's preferences, highlighting the importance of recommended decay times for listening rooms [42] (RQ3), and the subsequent standardization. In terms of the assessors' preferences, it was found that a critical value exists, above which the experience is reported as less pleasing. The work presented in Paper [D] further expands the research in sound reproduction in ordinary rooms, and depicts the need for perceptual control in such settings. The influence of the room in a sound reproduction scenario seems to be crucial, and the need for compensation schemes that address such degradation is eminent.

6.4 Limitations of the study

A number of limitations exists in this study relating to the methods, and the limited research focus. It is important to acknowledge these, identify the current pitfalls, and improve future studies.

In this work, we focused on certain applications, trying to capture the essence of sound in the limited context of domestic and automotive environments, when excited by a typical sound reproduction system. The need for a new perceptual assessment method was evident, which was then designed and implemented based on the contextual factors of this study. In consequence, a number of influencing factors was kept constant, though their effect is highly important for these phenomena in real life. That includes the type of the excitation signals, the source's properties, the relative position between source and listener, as well as perceptual and cognitive aspects, e.g., a multimodal experience. To better understand the effects of sound in domestic and automotive

---

[14] the first reflection points were kept constant in the within-room investigations. Naturally different rooms have different early reflection properties. The room's properties and early reflection points are summarized in Paper [D]

environments, we unequivocally need more studies in the domain. This work can only form a basis for further investigations in this field of research.

Further, SDM is a perceptually-based method. Although our results, as well as previous results in concert halls [36, 92] and critical listening environments [67] resemble what one could experience in real life, indicating adequate performance of the method, it is still unknown to what extent the physical field is accurately reproduced. In this work several steps were taken to ensure that the reproduced sound field matched the recorded one. This is possible in terms of pressure under ideal conditions[15]. However, the method, assumes that a point of the field resembles the true sound field. This is of course not a physically valid statement. Moreover, the method compromises the spatial distribution of reflections by assuming a discrete occurrence at each time point.

To explore these limitations, a series of physical analyses of the field was conducted. Figure 18 shows an example of the spatiotemporal analysis of an ordinary room, comparing the original sound field, as captured *in-situ* in the real room, and the reproduced sound field, that was presented to the subjects over a 40.3 loudspeaker array; this was the experimental apparatus used in the presented investigations. It can be seen that the simulated field is not accurately reproduced, especially in terms of its spatial representation (Fig. 18a). That is indeed an expected result, as the SDM discretize the space into a finite grid, fitting close-by energy to the nearest loudspeaker[16]. The magnitude response over the time and frequency, illustrate high similarity (Fig. 18b) between the two fields, as well as the *Impulse Response* (IR), shown in Fig. 18c.

The modal effects of the reproduction room, i.e., the anechoic chamber, are however apparent due to the chamber's restricted size. This can be observed in Fig. 18b as the lower frequencies ($< 120Hz$) of the simulated field are not reproduced accurately, especially after 20 ms. This is not an unexpected result, as the low frequency compensation applied at the listening position would only affect the direct sound.

The IR between the simulated and original fields (Fig. 18c) share temporal and amplitude similarities, indicating the ability of the method to deliver these properties adequately. The simulated IR is notably less smooth, relating to the IR imposed on the sound field by the electroacoustical properties of the 43 reproduction loudspeakers.

Although the representation of the field is not identical to the original, the perceived sound field using this method has been characterized by all assessors as natural and believable, some of which identified rooms and systems they had experienced before [67]. SDM seems to provide a perceptually adequate

---

[15] SDM synthesis uses the pressure captured by a single microphone. The summation of the spatially distributed pressure values will result to perfect reconstruction during the synthesis, assuming ideal reproduction settings.

[16] Other techniques are possible, e.g. VBAP. Here, we followed nearest neighbor fitting.

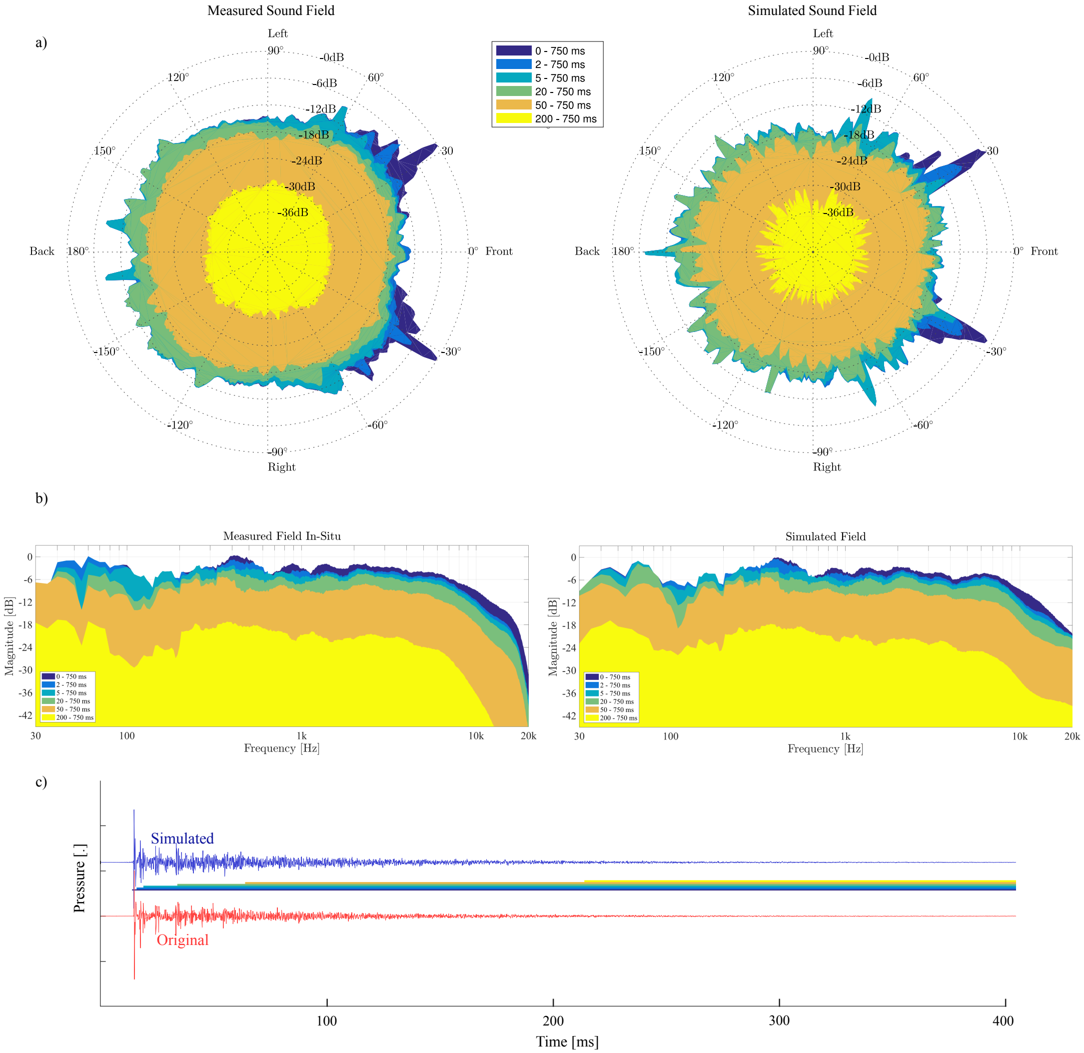


Fig. 18: Physical analysis of the measured sound field versus the reproduced, i.e., simulated, field in the laboratory. The measured sound field refers to the captured RIR using two loudspeakers in stereo configuration in an ordinary room, using identical apparatus as described in Paper [D]. The simulated field is the response of the 40.3 system, when the measured sound field is reproduced, i.e., what the listener would perceive during the experiment. At the top, (a) the spatiotemporal analysis [69] shows the energy distribution in space, in time intervals, where as (b) indicates these in terms of frequency. Finally, at the bottom (c) presents the summed Impulse Responses of the two fields, i.e. the original pressure measured in the room, and the pressure in the array when reproducing that field. NOTE: the original sound field was captured at 192k Hz, where as the simulated was captured at 48k Hz due to practical limitations. This explains the smoother response of spatiotemporal analysis (a) of the measured response compared to the simulated one, at a certain extend.

experience and an artefact-free representation of a 3D sound field[17]. Here, a systematic variation of room acoustics was followed and both the analysis and the synthesis of the fields matched our expectations, e.g., the physical results in Fig. 13 and the results in Paper [C].

To the author's best understanding the physical inaccuracies of the method are not well perceived by the human listener, due to the limited spatial acuity of the auditory system when exposed to complex sound fields. It is unlikely that the hearing system would attempt to spatially distinguish discrete reflections that arrive in succession and coherence [45]. In fact, the hearing system would follow *Gestalt* principles in such fields, in order to group several inputs into meaningful and known auditory objects. This may relate to a perceptual advantage of SDM compared to other spatial synthesis schemes.

The method's ability to deliver uncolored experience to the human listener, that is perceived as natural, may relate to Thiele's Association Model [21]. As the sound field is synthesized using SDM and reproduced over a grid of loudspeakers the auditory system is able to recognize the perceived experience and associate it with a real sound field. This is achieved by segregating the auditory objects and combining the signals from all the reproduction loudspeakers a single acoustical field. In consequence, although in practice a number of reproduction loudspeakers provide sound individually, their partial contribution is perceptually merged into a complex sound field, rather than being identified as discrete sources. This may highlight the ability of the SDM synthesis to provide coherent signals to all loudspeakers simultaneously, so that the human brain decodes these as a complete reverberant field.

6.5 Future Directions

The effects of sound in enclosures and reverberation comprise of boundless research streams and several phenomena that are not well understood. Using this work, its results, and applications as a basis, future work can expand in a number of directions. The exploratory approaches followed in the case studies could be expanded in a more systematic evaluation of sound fields, where certain aspects of the fields are studied in detail. The evoked sensations could then be contrasted to physical quantifiable metrics. For example, investigations in car cabin acoustics could focus on the perceived effects relating to source-listener positions, seat-to-seat variation, and human occupancy, to name a few.

In the work presented here, it was evident that the spatial distribution of reflections has a significant effect on the perceived sound experience. That is known to be influenced by the properties of the source. With current advancements in loudspeaker technologies, the interaction between loudspeakers' directivity and enclosures could be explored. It is likely that the perceived

---

[17] Perceptually artefact-free as no SDM studies have reported any attributes relating to technical issues.

properties of a room, i.e., its acoustics, could be perceptually controlled by such approaches, allowing a direct relation between a physical and highly controlled feature, and the perceived experience at the listening position.

Additional studies should be conducted to further validate the methods described in this work. The studies may focus on quantifying the limitations of these protocols. It is understood that such approach in perceptual terms is indeed not straightforward. To achieve this, the spatial reproduction of an acoustical field is required to be comparatively evaluated to the real field, by human assessors. Such approach is impractical and non trivial. One could employ binaural recordings in both the real field and the simulated one, and comparatively evaluate the perceived differences. Although such approach may involve other challenges, e.g., lack of spatially externalized fields, it allows a simultaneous and blind comparison of the fields in a relative manner, under the same evaluation protocol.

References

[1] B. Blesser and L. Salter, *Spaces Speak Are you Listening* (MIT Press),22-23 (2007).

[2] Webster, Douglas B. and Popper, Arthur N. and Fay, Richard R., *Evolutionary Perspectives – Invertebrates – Aspects of Hearing Among Vertebrates – Anamniotes – Non- Mammalian Amniotes – Mammals – Epilogue.* (Springer: New York), (1992).

[3] W. C. Sabine. *The Collected Papers on Acoustics*, volume 33. University of California, Los Angeles, 1922.

[4] L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York) (2004), pp. 1–653.

[5] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes", J. Acoust. Soc. Am. **130**, 835–49 (2011).

[6] T. Lokki, J. Patynen, A. Kuusinen, and S. Tervo. Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *Journal of the Acoustical Society of America*, 132(5):3148–3161, 2012.

[7] M. Barron, "Subjective study of British symphony concert halls", Acta Acust. united with Acust. **66**(1), 1–14, (1988).

[8] T. Hidaka and L. Beranek. Objective and subjective evaluations of twenty-three opera houses in Europe, Japan, and the Americas. *The Journal of the Acoustical Society of America*, 107(1):368–83, Jan. 2000.

[9] A. Gade. Acoustics in halls for speech and music. *Springer handbook of acoustics*, 2007.

[10] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, “Perception of Reverberation in Small Rooms : A Literature Study,” in *Proc. 55th Int. Conf. Audio Eng. Soc. Spat. Audio*, Helsinki, Finland, 2014.

[11] H. Kuttruff. *Room acoustics.* Spon Press, fourth edition, 2009.

[12] S. Bech, “Perception of Timbre of Reproduced Sound in Small Rooms: Influence of Room and Loudspeaker Position”, J. Audio Eng. Soc **42**(12), 999–1007 (1994).

[13] F. E. Toole, *Sound Reproduction: Loudspeakers and Rooms* (Focal Press, Boston) (2008), 1–550.

[14] Nielsen, *Music 360*, (2014).

[15] B. C. J. Moore. *An Introduction to the Psychology of Hearing.* Emerald, sixth edition, 2013.

[16] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, method and application* (Wiley) (2006), 1–442.

[17] S. Bech, R. Hamberg, M. Nijenhuis, K. Teunissen, H. Looren de Jong, P. Houben, and S. K. Pramanik. Rapid perceptual image description (RaPID) method. In B. E. Rogowitz and J. P. Allebach, editors, *Electronic Imaging: Science & Technology*, pages 317–328. International Society for Optics and Photonics, Apr. 1996.

[18] T. H. Pedersen and C. Fog. Optimisation of Perceived Product Quality. In *Euronoise II*, pages 633–638, 1998.

[19] M. Barron. Objective assessment of concert hall acoustics using Temporal Energy Analysis. *Applied Acoustics*, 74(7):936–944, July 2013.

[20] L. Cremer. *Die wissenschaftlichen Grundlagen der Raumakustik: Band I, Geometrische Raumakustik.* Hirzel-Velgard, Stuttgard, 1948.

[21] R. Thiele. Directional Distribution of Reflections in Rooms (In German). *Acustica*, 3(1):291–302, 1953.

[22] A. Marshall. A note on the importance of room cross-section in concert halls. *Journal of Sound and Vibration*, 1967.

[23] International Organization for Standardization, “ISO 3382-1: Acoustics - Measurement of room acoustic parameters, Part 1: Performance Spaces”, (2009).

[24] International Organization for Standardization, “ISO 3382-2: Acoustics - Measurement of room acoustic parameters, Part 2: Reverberation time in ordinary rooms”, (2008).

[25] L. Beranek. Concert hall acoustics—2008. *Journal of the Audio Engineering Society*, 56(7), 2008.

[26] A. H. Marshall and M. Barron. Spatial responsiveness in concert halls and the origins of spatial impression. *Applied Acoustics*, 62(2):91–108, Feb. 2001.

[27] J. Bradley. Review of objective room acoustics measures and future needs. *Applied Acoustics*, 72(10):713–720, Oct. 2011.

[28] T. Lokki. Throw away that standard and Listen: your two ears work better. In *International Symposium on Room Acoustics*, Toronto, Canada, 2013.

Ando, Y., and Imamura, M. 1979. Subjective preference tests for sound fields in concert halls simulated by the aid of a computer. Journal of Sound and Vibration, 65(2), 229–239.

[29] Y. Ando, M. Imamura, Subjective preference tests for sound fields in concert halls simulated by the aid of a computer. *Journal of Sound and Vibration*, 65(2), 229–239, 1979.

[30] A. Kuusinen, *Multidimensional perception of concert hall acoustics Studies with the loudspeaker orchestra.* PhD Thesis, Aalto University, 2016.

[31] S. E. Olive, P. L. Schuck, S. L. Sally, and M. Bonneville, “The Variability of Loudspeaker Sound Quality Among Four Domestic-Sized Rooms”, in *the 99th Conv. Audio Eng. Soc.*, October, 6–9, 1995, New York, USA, (Audio Engineering Society, New York), Prepr. No 4092.

[32] S. Bech, “Timbral aspects of reproduced sound in small rooms. I”, J. Acoust. Soc. Am. **97**, 1717–26 (1995).

[33] F. Christensen, G. Martin, M. Lydolf, M. Pauli, S. Woo-Keun, and P. Benjamin, “A Listening Test System for Automotive Audio - Part 1 : System Description”, in *the 118th Conv. Audio Eng Soc.*, May, 28–31, 2005, Barcelona, Spain, (Audio Engineering Society, New York), Prepr. No 6358.

[34] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S.H. Jensen, “A Rapid Sensory Analysis Method for Perceptual Assessment of Automotive Audio”, J. Audio Eng. Soc. **65**(1/2), (2017).

[35] F. Rumsey. Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society*, pages 651–666, 2002.

[36] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, “Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences”, J. Acoust. Soc. Am. **140**(1), pp. 551-562, (2016).

[37] International Telecommunication Union, “BS.1116-1: Methods For the Subjective Assessment of Small impairments in Audio Systems Including Multichannel Sound System”,ITU-R:Geneva, Switzerland, (1997).

[38] International Telecommunication Union, “Rec. BS.1534-2 Method for the subjective assessment of intermediate quality level of audio systems”, ITU-R:Geneva, Switzerland, (2014).

[39] C. C. J. M. Hak and R. H. C. Wenmaekers, “The impact of sound control room acoustics on the perceived acoustics of a diffuse field recording”, WSEAS Trans. Signal Process. **6**, 175–185 (2010).

[40] W. Evans, *The Influence of Loudspeaker Directivity upon the Perception of Reproduced Sound in Domestic Listening Rooms.* PhD Thesis, University of Surrey, 2012.

[41] International Telecommunication Union, “Rec.775-3 Multichannel stereophonic sound system with and without accompanying picture”, ITU-R:Geneva, Switzerland, (2012).

[42] International Electrotechnical Commission, “IEC:60268, Sound System Equipment— Part 13: Listening Tests on Loudspeakers”, IEC:Geneva, Switzerland, (1998).

[43] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse. Comparative Study of European Concert halls: correlation of Subjective preference with Geometric and Acoustic Parameters. *Journal of the Acoustical Society of America*, 56:1195–1201, 1974.

[44] R. K. Clifton, R. L. Freyman, and J. Meo, “What the precedence effect tells us about room acoustics”, Percept. & Psychophys., **64**(180), 180–188 (2002), doi:10.3758/BF03195784 .

[45] J. Blauert. *Spatial hearing: the psychophysics of human sound localization.* MIT Press, Cambridge, revised ed edition, 1997.

[46] S. Hameed, J. Pakarinen, K. Valde, and V. Pulkki, “Psychoacoustic Cues in Room Size Perception”, in *Proc. 116th Conv. Aud. Eng. Soc.*, May, 8-11, 2005, New York, USA, (Audio Engineering Society, New York), Prepr. No 6084.

[47] L. Savioja, P. Svensson, “Overview of geometrical room acoustic modeling techniques”, J. Acoust. Soc. Am., **138**(2), 708-730 (2015)

[48] C. Pike, R. Mason, and T. Brookes, "The effect of auditory memory on the perception of timbre", in *the 136th Conv. Audio Eng. Soc.*, April, 26-29, 2014, Berlin, Germany, (Audio Engineering Society, New York), Prepr. No 9028.

[49] M. Sams, R. Hari, J. Rif, and J. Knuutila, "The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence", J. Cogn. Neurosci. **5**(3), 363–368 (1993).

[50] V. Hulusic, C. Harvey, K. Debattista, N. Tsingos, S. Walker, D. Howard, and A. Chalmers. Acoustic Rendering and Auditory-Visual Cross-Modal Perception and Interaction. *Computer Graphics Forum*, 31(1):102–131, Feb. 2012.

[51] M. Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer-Verlag,2008.

[52] D. Botteldooren. Finite-difference time-domain simulation of low-frequency room acoustic problems. *J. Acou. Soc. of Am.* , vol. 98, pp. 3302–3308, 1995.

[53] W. Evans, *Binaural Resynthesis of Acoustical Environments. Technology and Perceptual Evaluation.*. PhD Thesis,TU-Berlin, 2014.

[54] N. Harris, "A comparison of modelling techniques for small acoustic spaces such as car cabins," in *122nd Conv. Audio Eng. Soc.*, Vienna, Austria, 2007, convention paper #7146.

[55] E. Granier, M. Kleiner, B. Dalenbäck, and P. Svensson, "Experimental Auralization of Car Audio Installations," *J. Audio Eng. Soc.*, vol. 44, no. 10, pp. 835–849 (1996).

[56] J. Blauert. *The Technology of Binaural Listening* . Springer:Berlin Heidelberg, 2013.

[57] S. E. Olive and T. Welti, "Validation of a Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems," in *Proc. 36th Int. Conf. Audio Eng. Soc. Automot. Audio*, 2009, pp. 1–7.

[58] R. Nicol, *Binaural Technology.* AES Monograph. New York: Audio Engineering Society Inc, 2010.

[59] P. Hegarty, S. Choisel, and S. Bech, "A listening test system for automotive audio - Part 3: Comparison of attribute ratings made in a vehicle with those made using an auralization system", in *the 123rd Conv. Audio Eng. Soc.* ,New York, USA, October, 5-8, 2007), (Audio Engineering Society, New York), Prepr. No 7224.

[60] W. L. Martens, H. Sakanashi, and W. Woszczyk, “Whole-body vibration associated with low-frequency audio reproduction influences preferred equalization,” in *Proc. 36th Int. Conf. Audio Eng. Soc. Automot. Audio*, Dearborn, USA, 2009.

[61] S. Germain, S. E. Olive, and T. Welti, “The Effect of Whole-body Vibrations on Preferred Bass Equalization of Automotive Audio Systems,” in *127th Conv. Audio Eng. Soc.*, New York, USA, 2009, convention paper #7956.

[62] H. Wittek. Perceptual differences between wavefield synthesis and stereophony. 2007.

[63] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, “Spatial Sound With Loudspeakers and Its Perception : A Review of the Current State”, Proc. IEEE **101**, pp. 1920–1938, (2013), doi:10.1109/JPROC.2013.2264784 .

[64] V. Pulkki, Spatial sound reproduction with directional audio coding. *J. Aud. Eng. Soc.*, 55(6), pp 503-516, 2007.

[65] J. Merimaa, V. Pulkki, Spatial impulse response rendering I: Analysis and synthesis. *J. Aud. Eng. Soc.*, 53(12), pp 1115-1127, 2005.

[66] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, “Spatial decomposition method for room impulse responses”, J. Audio Eng. Soc. **61**(1/2), (2013).

[67] S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki, “Preferences of critical listening environments among sound engineers”, J. Audio Eng. Soc. **62**(5), pp. 300–314, (2014).

[68] J. Pätynen, S. Tervo, and T. Lokki, “Amplitude panning decreases spectral brightness with concert hall auralizations”, in *55th Int. Conf. Audio Eng. Soc. Spat. Audio* ,August 24–27, 2014, Helsinki, Finland, (Audio Engineering Society, New York).

[69] J. Pätynen, S. Tervo, and T. Lokki, “Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses.”, J. Acoust. Soc. Am. **133**, pp. 842–57 (2013).

[70] S. Tervo, J. Pätynen, N. Kaplanis, M. Lydolf, S. Bech, and T. Lokki, “Spatial Analysis and Synthesis of Car Audio System and Car-Cabin Acoustics with a Compact Microphone Array”, J. Audio Eng. Soc **63**(11), pp. 914–925 (2015), http://dx.doi.org/10.17743/jaes.2015.0080.

[71] G. Martin and S. Bech, “Attribute Identification and Quantification in Automotive Audio - Part 1: Introduction to the Descriptive Analysis Technique”, in *Prepr. 118th Conv. Audio Eng. Soc.* (Barcelona, Spain) (2005).

[72] H. T. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices* (Springer: NewYork), (1999) pp. 1–471.

[73] J. Delarue, B. J. Lawlor, and M. Rogeaux, *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research* (Woodhead Publishing, Elsevier: Cambridge, UK), (2014), pp. 119-148.

[74] P. Varela and G. Ares, *Novel Techniques in Sensory Characterization and Consumer Profiling* (CRC Press: Boca Raton, USA) (2014), pp. 175–206.

[75] V. Dairou and J.-M. Sieffermann, “A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile”, J. Food Sci. **67**, pp. 826–834 (2002).

[76] R.L. Robart and L.D. Rosenbaum, “Reflected sound helps your brain recognise the type of space:”, Hearing Space: Identifying rooms by reflected sound. In J.Heft & K.L. March (Eds.), Studies in perception and action XIII , (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.) (2004), pp.152-156.

[77] A. Tajadura-Jiménez, P. Larsson, A. Väljamäe, D. Västfjäll, and M. Kleiner. When room size matters: acoustic influences on emotional responses to sounds. *Emotion (Washington, D.C.)*, 10(3):416–22, June 2010.

[78] J. Sandvad. Auditory perception of reverberant surroundings. *The Journal of the Acoustical Society of America*, 105(2):1193, Feb. 1999.

[79] P. Strumillo, M. Pelcynski,M. Bujcz, “Space Perception by Means of Acoustic Images: an electronic travel aid for the blind,”, in *ACOUSTICS 33th International Acoustical Conference – EAA, Slovakia.* (Slovakia) (2011).

[80] E. Calcagno, E. Abregu, M. Eguía, and R. Vergara. The role of vision in auditory distance perception. *Perception*, 2012.

[81] P. Larsson, D.Västfjäll, and M. Kleiner, “Multimodal interaction in real and virtual concert halls”, J. Acoust. Soc. Am. **115**, pp. 2403 (2004).

[82] P. Larsson, and D. Västfjäll, “When what you hear is what you see: Presence and auditory-visual integration in virtual environments”, in *Proc. 10th Annu. Int. Work. Presence*, 11–18 (Barcelona, Spain) (2007).

[83] D.L. Valente, J. Braasch, “Subjective expectation adjustments of early-to-late reverberant energy ratio and reverberation time to match visual environmental cues of a musical performance”, Acta Acust United Acust **94**,6, pp. 840-55 (2008).

[84] D. L. Valente, J. Braasch, and S. a. Myrbeck. Comparing perceived auditory width to the visual image of a performing ensemble in contrasting bi-modal environments. *The Journal of the Acoustical Society of America*, 131(1):205–17, Jan. 2012.

[85] J. Schnupp, I. Nelken and A. King, *Auditory neuroscience : making sense of sound* (MIT Press: Boston) (2011), pp. 1–356.

[86] P. Strumillo Electronic interfaces aiding the visually impaired in environmental access, mobility and navigation. *3rd Int. Conf. Hum. Syst. Interact. HSI'2010 - Conf. Proc.*, pp. 17–24, 2010.

[87] M. Bitner. Servicescapes: The impact of physical surroundings on customers and employees. *Journal of marketing*, 1992.

[88] S. Kemmis,R. McTaggart and J. Retallik, "The Action Research Planner." (Karachi: Aga Khan University Institute for Educational Development) (2004).

[89] T. H. Pedersen and N. Zacharov, "The development of a Sound Wheel for Reproduced Sound", in *the 138th Conv. Audio Eng. Soc.* , (May, 7-10, 2015),Warsaw, Poland, (Audio Engineering Society, New York), Prepr. No 9310.

[90] N. Zacharov, and T. H. Pedersen. "Spatial sound attributes - development of a common lexicon", in *the 139th Conv. Audio Eng. Soc.* , (Oct., 21-1, 2015),New York, USA, (Audio Engineering Society, New York), Prepr. No 9436.

[91] A. Lindau, V. Erbes, S. Lepa, H.J. Maempel, F. Brinkman, and S. Weinzierl, "A spatial audio quality inventory for virtual acoustic environments (SAQI)", Acta Acust. united with Acust. **100**, pp. 984–994 (2014).

[92] J. Pätynen, T. Lokki, "Concert halls with strong and lateral sound increase the emotional impact of orchestra music ", J. Acoust. Soc. Am. **139**, 1214–1224 (2016).

[93] G. Thiele, *On the localization in the superimposed field.* PhD Thesis, TU-Berlin, 1980 (Translated by Neher).

Part II

Papers

Paper A

Perception of Reverberation in Small Rooms: A Literature Study

Neofytos Kaplanis, Søren Bech, Søren Holdt Jensen, and Toon van Waterschoot.

The paper was published in the
Proceedings of 55th International Conference of Audio Engineering Society,
on Spatial Audio , Helsinki, Finland, August, 2014.

The layout has been revised.

Abstract

Reverberation is considered as one of the fundamental perceived properties of an acoustical space. Literature is available on the topic and currently a range of sciences have contributed in understanding the properties of reverberant sound fields and the relevant auditory processes. This paper summarises the current literature following a top-down approach. It identifies the perceptual aspects of reverberation and attempts to establish links to physical measures, focusing on small rooms. Results indicate that the current acoustical metrics often have limited correlation to the perceptual attributes of reverberation and conclusive measurement data is restricted, especially for small spaces. A proposal for perceptually-based experiments is presented, aiming to further understand the links between physical properties of rooms and their effects on perception.

1 Introduction

In enclosed spaces, the interaction between the sound source and the room's boundaries produce a distinctive sound field, which is commonly characterised as *reverberation*. Reverberation is normally used as a global umbrella to describe a set of physical, perceptual and affective features of certain sound fields. In the physical domain, a number of objective metrics describing the acoustic properties of reverberant sound fields have been established and standardised [1–3]. However, what describes reverberation in the perceptual and affective domains -that is the human perception— is yet, not fully understood [4, 5]. This paper examines and analyses the published scientific literature related to reverberation, in an attempt to identify the most important perceptual characteristics of reverberant sound fields (Section 2), and seeks their relations to the proposed physical metrics (Section 3). This approach will form the benchmark framework for further investigation on the properties of reverberation aiming towards the perceptual control of reverberant sound fields in typical-sized, domestic spaces (Section 4).

As the direct investigation of the perceptual aspects of reverberation is very limited for small rooms, it is worthwhile to evaluate the scientific literature in concert hall acoustics, spatial audio, psychoacoustics, and acoustic measurement in an attempt to merge findings, identify common paths, and seek the applicability of these relationships in every-day listening spaces. Here, the literature is examined and analysed based on the principles of the *Filter Model* [6–8]; the content of the paper will be categorised based on the two domains of the model, the physical and the perceptual. A brief introduction to the model is given below.

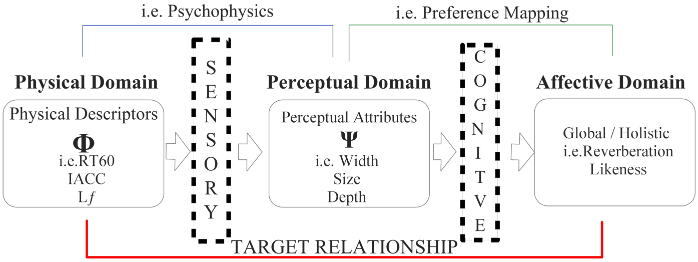


Fig. A.1: The Filter Model, comprising three domains, physical, perceptual, and affective separated by a sensory and a cognitive filter respectively.

1.1 Filter Model

According to the filter model (Fig.1), a *physical stimulus* (e.g. sound event) is perceived after being filtered by our senses (e.g. hearing) resulting to a *perceptual entity* (e.g. auditory event). The percept also passes through a *cognitive* filter, which represents the non-sensory factors such as mood, expectations, and memory. One advantage of the filter model is the ability to estimate the human *affective* response based on physical measurements, by using the perceptual domain as the bonding element. This approach effectively associates these domains by dividing it into two steps. The identification of: (1) links between the physical measurement and perceptual attributes — i.e. psychophysics and perceptual models, and (2) the links between perceptual attributes and the affective response for a given global percept — i.e. preference mapping.

1.2 Physical Domain - Reverberation

The physical basis of reverberation is briefly presented below, as several metrics reviewed in this manuscript are based on these principles. However, the physical description of a reverberant sound field is not the focus of this paper, and the interested reader is referred to [9, 10].

The Reverberant Sound Field

Reverberation was primarily thought to be a global acoustic behaviour of an environment: the distinct prolonging sound caused by reflective surfaces and slow speed of sound [9, 11, 12]; resembling what humans sense as *Reverberance*. Through years of research it was clear that the perceived sound field, as seen from the listener's position, consist of two distinct stages. Figure 2 shows a simplified example of a RIR. Sound typically travels along the direct path - towards the receiver - and it arrives after a propagation delay *t*. Then, *early reflections* originating from nearby objects/boundaries (i.e. ceiling, floor, sidewalls) arrive at the receiver. As the reflections spread in a finite space and speed, they interact with each other, effectively increasing the echo density over time and decreasing their intensity. This interaction results to another typical pattern of an RIR, referred to as *late reflections*.

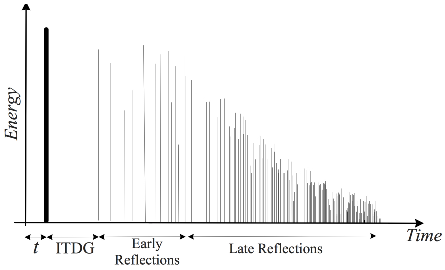


Fig. A.2: A typical Impulse Response in a room.

It should be noted that in this paper 'reverberation' will be used as the global descriptor of the acoustic behaviour of the room, unless otherwise stated. The term 'Reverberance' will be used as a perceptual descriptor of sensing the late part of reverberation. In the following text, the current objective metrics used to characterise reverberant sound fields will be presented, followed by the perceptual attributes found in similar literature.

Reverberation's Objective Metrics

Research in acoustics, primarily in concert halls, initially focussed on the objective and physical characterisation of reverberation. Numerous physical aspects and mathematical explanations of reverberant fields have been investigated and their findings established common acoustic measurements such as *Reverberation Time* (RT), EDT, G, *Early Energy Fraction* (J_{LF}), *Late Lateral Sound Level* (J_j), IACC, $C_{50/80}$ and others [1, 3, 13]. These metrics target the quantification of the average physical properties of the reverberant sound field in concert halls, 'providing a single number, which is relevant to at least some aspect of the acoustical quality' [14], p189. Recent studies [5, 15–17, 68] however, challenge the perceptual relevance of the established metrics while others identify uncertainties, imprecision, inadequate frequency range, and errors in acquiring and computing these parameters [5, 18–25].

It seems therefore central to revisit these relationships and attempt to define the major perceptual attributes of reverberation *per se*. By establishing the perceptual attributes of reverberation, the direct investigation on the perceptual relevance of the current and/or novel metrics could be performed following a top-down approach, keeping the human perception as the core element of this analysis. A stronger link between physical metrics and perceptual attributes will enable a direct association between the physical properties of a space, objective metrics, and the final piece of the audio chain, the human listener. In the next section, the perceptual relevance of reverberation is examined, and the most frequently elicited perceptual attributes found in the literature will be discussed.

1.3 Perceptual Domain - Auditory Perception

In search of capturing the essence of reverberation, fundamental work in concert hall acoustics [26] revealed the importance of the time required for the sound pressure level to decrease by 60dB from its initial level, the RT_{60}. RT_{60} was related to the global perceptual attribute of Reverberance, a key indicator of acoustical quality. This led to an ‘optimum value’ of RT [27]. Nevertheless, over the last decades acousticians recognised that achieving perceptually “good acoustics” was more than reaching an optimum RT. Lothar Cremer [28] illustrated that reflections’ properties hold high importance in reverberant fields, such as their series of arrival, density, and their global decay characteristics. Similarly, many studies supported this notion [29, 30] relating the early part of reverberation to spatial characteristics of the field [31, 32], and more recently [33] to even enhanced dynamic range of an orchestra. Following these findings, acoustic research focused on understanding the perceptual mechanisms of reverberation and its parameters, before addressing their physical characteristics (see [5, 34, 35]). In these investigations several methods have been applied to evaluate the perceptual aspects of reverberation and their properties. The major studies have conducted *in-situ* evaluations of concert halls using questionnaires [36–41], laboratory-based studies intending to reproduce hall acoustics [42–48], while others focussed on the recording and reproduction techniques for reverberant and spatial fields [49–55].

Within the current literature there is a general agreement that reverberation is not a one-dimensional phenomenon, but it is rather related to a set of perceptual attributes all of which contribute to the acoustical identity of a space (see [3, 36, 56, 57]). It is therefore understood that the perception of the global acoustic behaviour of a room, namely its reverberation, is influenced by certain perceptual attributes such as: the source’s apparent dimensions *Depth, Distance, Width, Size*, as well as *Timbre* [58], *Loudness* [59], *Distance* [59–61], *Perceived Room Size* [62], *Clarity, Localisation-Diffuseness, Transparency*, as well as semi-abstract ‘immersive’ percepts [51], such as *Warmth, Intimacy-Presence, Fullness, Spaciousness, Envelopment, Reverberance* and many others.

2 Perceptual Attributes of Reverberation

This decomposition of the holistic experience of reverberation into several attributes (i.e. Size, Timbre), can be thought as the process of visual face recognition [57]. A face can be perceived as a holistic entity i.e. a person's face, while it can be decomposed into individual attributes i.e. his/her eye colour. The advantage of looking into individual attributes (i.e. eyes color) rather than the global image (i.e. face) is that the perceptual aspects of the global and complex phenomenon can be justified objectively even when a small sample of human subjects is questioned. In this section, a non-exhaustive but representative summary of the most salient features of reverberation is presented (see Figure 3) based on the frequency of occurence that they appear in the literature reviewed. Following our research questions, the practical limitations, and the scope of application for domestic spaces, attributes related to general quality (i.e. Naturalness), attributes that are multidimensional by nature (i.e Timbre) as well as technical-related attributes (i.e. Distortion) that are depicted in Figure 3, will not be discussed further. The attributes of interest will be presented below following a cohesive order to aid the ease of reading, starting from high level space-related attributes to lower level source-related attributes (see [51]), rather than their frequency of occurence.

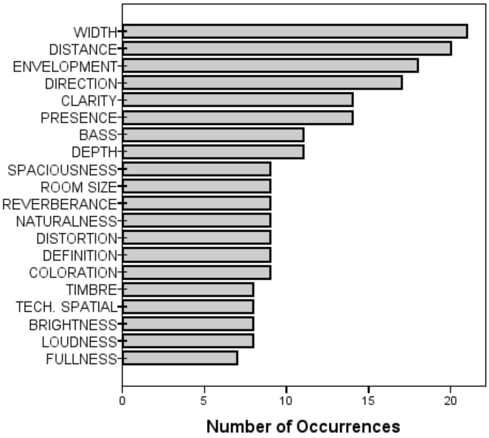


Fig. A.3: Histogram depicting the rank-order of 328 perceptual attributes found in elicitation studies which were categorised in 74 unique terms.

2.1 Room Size

The apparent Room Size is fundamental in identifying the space we are in; a psychological process known as *space perception* [63, 64]. The perceived room size is believed to modulate our perception, cognition [65] and set our mental state, even without us being aware i.e. when sensory input matches expectations [13]. Room Size has been identified as a significant attribute in spatial acoustics and standardised as a perceptual rating in the multidimensional evaluation of spatial reproduction [66]. Recently [41] it was also found to be a mutual attribute in describing acoustic spaces by both acousticians and musicians. The direct relationship between perceived Room Size and Reverberance is well deceptive in the literature (see [62]). This link is also supported by theoretical acoustics [9, 26], measurements [26, 67], and perceptual studies [68]. Rumsey [51] noted that the perception of Room Size does not require strong sense of Presence. In addition, Lokki *et al.* [4] suggested that Envelopment and Room Size are elicited independently, explicitly disassociating the 'sense of being enveloped' in a sound field and the perception of Room Size. Similarly subjects were able to distinguish between the perception of Room Size and Room Width and judge them individually [50].

Moreover, the *Direct-to-Reverb ratio* (D/R) is claimed to provide Distance cues, though individual differences are clearly apparent when using this cue. This variability is attributed to listener's mistake of using loudness as the main cue, while it is not a salient cue for Room Size [62, 69]. The majority of the studies reviewed above, credit RT as the main cue for apparent room size. Yet, the relationship between RT and increased physical size is not linear as it has been demonstrated that the auditory perceived Room Size differs from the physical, volumetric size [63, 70–72]. For example it is common to perceive an empty room bigger than what it really is.

Following the literature, it can be argued that the physical size of the room [71], the Acoustic Support (ST) originating by early reflections [60, 67] and the related decay of the reverberant field (RT) have the most influence on the apparent room size. However, the relationship is not linear [71], thus RT should not be taken as the *Lethe* to perceived size. In fact, RT changes in small rooms revealed no perceptual effect [67], indicating a possible influence of earlier reflections. Further, it could be seen that the auditory system evaluates the size of a room using different psychoacoustics mechanisms for different rooms [67]; an equivalent to perceived *distance parallax* (see [73]). It is therefore argued that room perception is not simple and linear as the cross-modal interaction plays a vital role in this process, by even violating classical acoustic theories.

2.2 Room Presence-Intimacy

Presence has been recognised as the perceptual sense of being inside an enclosed space and feeling its boundaries [51] - a hearing-equivalent of 'seeing' the walls of a room [36]. 'Presence', used mainly by recording/broadcasting engineers, is thought to be equivalent to 'Intimacy' used in concert hall acoustics [36, 51, 74]. In this paper only the term Presence will be used.

Since the early investigations on listener's preferences in concert halls [36, 37], Presence has been credited 40% of the overall quality in halls; compared to 15% attributed on Reverberance [13, 34]. In fact, Letowski [75] includes Presence as a major component of sound quality assessment. Following the definitions given in each study reviewed here, it is clear that perceptual attributes such as *sense of space* [55, 76], *feeling of space* [77], *feeling of Presence* [52], *Immersion* [34], and *Proximity* [40] all relate to the perceptual experience used to describe Presence. Sotiropoulou *et al.* attempted to link Proximity to Nearness and Envelopment in a single rating. However, it has been recently shown that that Proximity and Envelopment form separate factors when assessing concert halls [16]; similar findings were shown for Envelopment and Presence in sound reproduction assessment [51]. Presence has also been used to describe the *Naturalness* of a sound field [52, 67], an attribute that was later found to be strongly correlated with Presence [50]. In fact, in the domain of virtual environments, Presence and Immersion were sometimes treated as equal. However, it has been argued that Immersion requires a self-representation in a virtual environment, whereas Presence is a state of consciousness, a more high-level psychological sense of being somewhere [78, 79].

The main contributor towards the perception of Presence is thought to be the time period between the direct sound and the first reflection (see Figure 2), which is referred to as *Initial Time Delay Gap* (ITDG). The optimal ITDG towards intimate environment is believed to be around 25ms in concert halls [36], or 20ms in operas [80]. An alternative measurement known as Time-Delay Spectrometry [81] was also used to investigate effects of ITDG in smaller rooms (recording studios). These studies [82, 83] related ITDG with comb-filtering effects due to early reflections summing.

Overall, Presence seems to be an important attribute contributing to the 'sense of space' as a natural habitat, in virtual environments, as well as being a pre-requisite of Envelopment in an acoustic space [51]. Still, the robustness of the current metrics (e.g. ITDG) is not well established, and standardised methods are not available. The authors would therefore like to motivate investigations towards a more sophisticated and robust method in calculating ITDG. For example, techniques used in signal processing for radar and sonar applications such as *Time-Delay Estimation* [84] (see also [85]), could be employed. This may enable a direct evaluation of ITDG thresholds and perceived attributes, such as Presence/Intimacy, based on objective metrics.

2.3 Spaciousness

In this paper, 'Spaciousness' is defined as the perceptual sense of feeling immersed in an acoustic space, and the experience of being enveloped by its reverberant sound field [57, 75, 86–89]. This sensation has been also expressed as equal to or a part of *Spatial Impression* [44, 46, 52, 90–92], and *Spatial Responsiveness* [30, 93], which were all found to relate to the listeners' preference of concert halls and their perceived qualities [13, 36, 43, 55, 89, 91, 94–99], as well as quality of sound *per se* [75, 100]. These immersive attributes are usually used as global, cover-it-all terms, and they cannot form a one-dimensional, well-defined perceptual attribute that is needed in practice [51]. Still, if the linguistic and cultural ambiguities are omitted, it could be argued that most of these investigations aimed to describe two perceptual experiences: (a) a modification of the perceived dimensions (Width/Size) of the sound source, and (b) the sense of being enveloped and surrounded by the reverberant sound field. In fact, Morimoto and Maekawa [90] demonstrated that Spaciousness (Spatial Impression in their words) comprises of two major components. These components were later identified as separable percepts attributed to the *Apparent Source Width* (ASW) and *Listener Envelopment* (LEV) [46], and standardised [1] as components of Spaciousness (see [101]).

There is a common understanding in the literature that Spaciousness is primarily related to lateral reflections, hence, influenced by reflection's properties such as time, level, angle of incidence, spectrum, as well as the total sound level of both reflections and direct sound (see [91, 93, 102]). Consequently, it is argued that Spaciousness is influenced by fluctuations of *Interaural Time Difference* (ITD) and *Interaural Level Difference* (ILD) over time [87, 92, 103–107], see also [17]. In addition *Binaural Quality Index* (BQI) [36, 108], a physical metric based on IACC, was found to relate linearly to Spaciousness [109, 110]. Okano, Beranek and Hidaka [101] demonstrated that specific frequency bands make different contributions to aural Spaciousness.

Nevertheless, the reviewed literature contends that Spaciousness should be abbreviated as the global percept of feeling immersed, comprising of: a surrounding impression (LEV) and a broadness effect (ASW) [51, 111]. Thus, equating Spaciousness to a single aspect i.e. Width/Broadness of a source [112, 113] should be avoided (see [87]). The specific properties of LEV and ASW will be discussed below.

2.4 Listener's Envelopment - LEV

Listener's Envelopment (LEV) has been outlined in the literature as the perceptual sense of feeling in the centre of- [50] and surrounded by- [101] a reverberant sound field; as the 'analogous of swimming underwater than being sprayed by a water hose' [57]. LEV has been associated with the *Degree of Fullness* [99], *Room Impression*, the *extent of Immersion* [114], and *Immersion* (see [115]). Based on literature, one could identify a question that reappears in various papers: 'is envelopment a by-product of a very wide ASW?'. Similarly, Rumsey proposed [51] that envelopment should be classified as (1) *Environmental Envelopment* – following LEVs definition, and (2) *Source-related Envelopment* – describing the envelopment effect created by anechoic sound sources as found in sound reproduction studies (see [51]). Nevertheless, LEV seems to follow a combination of perceptual mechanisms, all of which contribute to a more global percept of feeling enveloped and surrounded within the acoustic space; an experience that is likely to increase immersion and preference in reverberant environments. The physical properties of reverberation that influence these percepts seem to be the spatial distribution of reflections [99], including front-back [116] and vertical directions [117], as well as the level of direct sound [62] and the overall sound level at the receiver position [46].

2.5 Apparent Source Width - ASW

One of the fundamental perceptual characteristics of a sound source in a room is its apparent *Width.* Width, or more commonly referred to as Apparent Source Width (ASW) [1], and Auditory Source Width [118] describes the perceived horizontal size of a sound source, and it has been defined as the spatially- and temporally- fused *auditory image* of the original sound and early reflections [101, 119]. The perceived Width has been identified as a major component of acoustical quality in several perceptual studies [16, 50, 60, 95, 120, 121] and it is also included in acoustical quality standards [66, 122] under its most common label of ASW. In fact, the rank-order analysis performed in this paper, depicts Width as the most commonly elicited attribute in reverberant and spatial sound fields. ASW has been associated to the sensations of *Broadness* [31], *Diffusion*, *Blurriness* [96] and ambiguity in source angular *Localization* [50] (see also [114]). Rumsey [51] argued that the signal properties elicit different perceptual experiences, which may explain the confusion in the literature i.e. the arbitrary definition of a 'source'. He then proposed a *scene-based* paradigm where he distinctly categorised width for (1) a single source -– the *Individual Source Width* (ISW) -– (2) for a group of cognitively similar sources -– the *Ensemble Width*, and the (3) *Environmental Width*, which the author linked to Spatial Impression (BSI), as defined by [87, 111], and the human ability to isolate the reverberant sound fields originating from the recording room and reproduction room, as individual entities; initially hypothesised by [123]. The physical parameters of the ASW have been primarily linked to early reflections' level, direction, frequency content, and their structure (see [31, 51]), total Sound Pressure Level [30, 108, 109, 124] as well as the source-receiver Distance [125]. Hidaka *et al.* [126] estimated the greatest effect on enhancing perceived ASW from reflections at angles of ±60°, while no effect was found due to the reflections originating from behind [116] or above the listener in both concert halls [117], and domestic listening rooms [127]. Similar to Spaciousness, the proposed ASW metrics primarily follow the binaural properties of the human auditory system in both temporal and frequency domains [17, 87, 92, 103, 106, 107, 111, 128, 129]. The most prominent metrics include IACC [1, 130, 131] and simile i.e. *Degree of Interaural Cross Correlation* (DICC) [132], BQI [23], and *Interaural Coherence* (ICC) [19, 103, 104]. Links were also established with newer metrics such as J_{LF} [1, 91, 108], G, combined J_{LF} and G known as *Degree of Source Broadening (DSB)* [13]. Moreover, ITD [133, 134] and loudness measures [135, 136] were also related to the perception of ASW. Still, the standardised estimations of the perceived ASW include several imperfections, resulting disengagement to the perceptual experience. These imperfections might be linked to the claimed frequency dependency of ASW [31, 91, 92, 137], as well as its influence by Distance which are not considered by these calculations [21, 125]. In fact it is argued that there is no reliable relationship between 1-IACC$_E$ or J$_{LF}$ with perceived ASW [5, 21, 125] in its current form.

Overall, it is evident that although ASW is a well-defined perceptual attribute describing the perceived dimensions of an auditory event, the current physical metrics encompass high level of uncertainties [21, 125], which challenge their perceptual relevance *per se.* The inclusion of human auditory processess in the estimations i.e. binaural processing (i.e. ITD, ILD [138]), frequency dependent Interaural Correlation ([139]), and the Precedence Effect may enable better estimation of the perceived ASW.

ASW Vs LEV

Although ASW and LEV have been identified as separate attributes, they both contribute in a more global percept, Spaciousness. Thus, it is argued [111] that it is unlikely to perceive a room as spacious (i.e. 'large' and 'open') if only one of the two attributes is elicited. Further, the arbitrary boundary point 80ms [1]-100ms [3] between early and late energy in halls, as well as practical issues with standardised methods (see [19]), point towards the need of more research in the area, especially when different sized rooms are considered. Perhaps, measures such as *Echo Density* [58] and *Centre Time (Ts)* [1] may indicate a more perceptually relevant temporal boundary between LEV and ASW.

2.6 Source Depth

Depth is a perspective-sense, which is identified mainly, but it is not limited to, reproduced sound [51]. Although, it has been considered as an elusive concept, as some listeners may have difficulties to perceive it, several studies suggest that during elicitation test, subjects described [52] and drew [140] sources as being curved and flat. Depth can be thought as a source's dimension, but concurrently the apparent Depth of the room may also elicit this sense. In fact, Rumsey [51] proposed an attribute known as Environmental Depth and attempted to relate Depth to Spaciousness -as defined by Griesinger [111]. The perception of Depth for a sound event has been attributed to frequency decorrelation [52, 140] and lateral reflections produced by components lower than 3KHz are responsible for the perception of Depth [103]. It has been also been noted that the directivity of a sound source influence the Depth of a sound even in a sound reproduction system, where directivities close to dipole have the biggest positive influence on perceived Depth [141].

2.7 Source Distance

Distance, is considered as an aspect of source localisation mechanism [142] a natural practice of our auditory system. In the literature attributes such as *Nearness* [77], *Presence* [51], *Closeness* [143], seem to have been used as linguistic alternatives of perceived Distance. It is noted that although reverberation normally degrades localisation abilities, the distance perception is believed to be enhanced [144]. In fact, room reflectivity [145, 146], and especially the early reflections play a major role in Distance perception [147], whilst the number of reflections [73], as well as the angle of incidence [124] influence this relationship. Auditory Distance perception has been also linked with D/R [61, 148] (when RT is high enough [149]), Spectrum (for familiar sounds [150, 151]), and binaural cues [152]. Moreover, it is argued [147] that loudness is not a major cue of Distance in typical rooms, as one would expect; this was also demonstrated in concert halls [153, 154], even under blindfold scenarios. This disconnection between perceived distance and the actual physical distance [147] could be linked to a loudness memory that humans seem to posses. This memory produces a crude estimation of the Distance if required [155, 156], even following only a visual mental reference [70] and other non-audio cues (i.e. vision, familiarity, expectation). It is therefore evident that humans use several cues [156] in Distance estimation, the weighting of each is highly adaptive (see [148]) and the related perceptual constructs seem multidimensional.

3 Discussion

The perceptual relevance of reverberation is highly apparent in the acoustics literature, and the current research focus has been shifted towards understanding perceptual aspects of reverberant sound fields rather than purely physical metrics. These studies highlight the importance of attributes related to early reflections, and more specifically the reflection pattern they exhibit [36, 60]. Still, perceptual characteristics of sound are difficult to epistomise and the lack of common vocabulary make it difficult to merge studies' results. Nevertheless, following the descriptions given in each study, this investigation revealed the most commonly elicited attributes in revebrant sound fields and several links to objective measurements have been presented (Section 2). A graphical summary is provided in Figure 4, depicting the perceptual attributes discussed in this manuscript and their relationships found in the literature. In the next paragraphs, the central topics are addressed, and the identified issues within this stream of acoustic research are discussed. The applicability of these findings in small rooms is reviewed (Section 3.1) and a proposal for perceptual-based experiments for reverberation in such rooms is presented (Sections 4-5), including considerations and points of attention.

3.1 Small Vs Large Rooms - Applicability

The physical differences between large rooms (i.e. auditorium) and small rooms (i.e. living room) will be discussed, as a tool to identify common paths and differences to our findings. These differences can be categorised by their three main causes: (1) the physical parameters of the propagation medium (i.e. the room), (2) the characteristics of the sound source used to excite the room, and (3) the psychological and psychophysical properties of the receiver, the human listener.

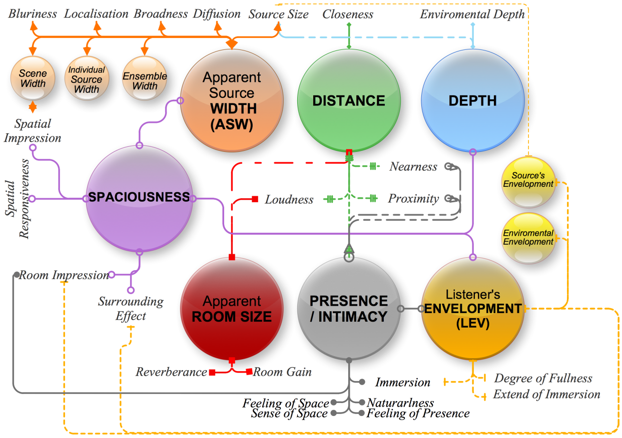


Fig. A.4: Paper overview based on the perceptual attributes presented. Large circles denote the fundamental attributes, while the small circles indicate their possible subcategories. Lines indicate the links between attributes as found in the literature.

The Medium - The Room

Due to lower volumetric dimensions, the typical RT in domestic spaces is much lower than a typical hall. These differences also influence the echo density of the reverberant soundfield in small rooms, exhibiting frequency irregularities and modal behaviour especially at Low Frequencies [10, 123]. The shorter lengths between boundaries in domestic rooms, impact the temporal distribution of reflections, introducing shorter ITDG, as well as stronger and distinct early reflections. Moreover, the first reflections typically originate from the floor and ceiling, rather than lateral directions as in concert halls (see [157]). Based on the reviewed perceptual attributes (Section 2) these characteristics are likely to influence the perception of Distance, ASW, Presence and apparent Room Size in smaller spaces. Moreover, it is expected that some objective metrics used in concert halls may not be suitable for smaller spaces. The temporal discrimination between early and late reflections at 50/80ms is not ideal, considering that most of the reflective energy in small rooms occurs within the first

50ms [111]. Thus, common measures that follow this strategy such as *Early Interaural Cross-Correlation* (IACC_E), BQI, and $\mathrm{C}_{50/80}$ may not indicate the appropriate measure of their related subjective impressions. The Centre Time (T_s) [1], a measure of the energy distribution over time, may serve as good measure for the physical boundary for early/late reflections.

The Source - Loudspeakers or Orchestra

In both halls and listening rooms, the scope is to experience music. However, a major difference between domestic rooms and concert halls is the sound source. In a hall the source is typically an orchestra based at a specified and acoustically designed area. In small rooms a reproduction system is involved, and the room's design requirements often discard acoustical performance. Consequently, an additional set of parameters is added which modify the perceived acoustical experience (for review see [123, 127]). Imperfections of the transducers, their placement, directivity, and performance of the system also influence the perceived experience (see [157]). Moreover, the perceived sound includes at least two distinct reverberant fields, that of the recording room, and also the reproduction one. Still, listeners are able to distinguish between the two independently [51, 123]. In conclusion, techniques used in concert halls, for example the use of a single omnidirectional loudspeaker to simulate the sound source, may not reveal perceptually relevant results.

The Receiver - The Human Listener

The surrounding environment is found to alter humans' cognitive processing (i.e. emotional state, alertness). Moreover, humans have certain expectations of the acoustics of the room [69] and have already memorised schemata to aid faster processing of complex scenarios; even for typical reproduction setups, such as 2ch. stereo [158]. This preconception may create inattentional blindness to auditory cues [159], as well as information suppression due to the already established space characteristics from other inputs such as vision [160, 161]. It seems therefore important to remove these biases in subjective evaluations of small room acoustics, for example by conducting the experiments in ecologically valid rooms, in matched visual and auditory environments, unfamiliar rooms, and/or blind setups.

3.2 Dealing with Perception

In order to tackle the issues related to reverberation, one needs to deal with physical acoustics as well as perceptual acoustics. The existence of a non-linear and highly adaptive processor within our auditory system, the brain, makes the establishment of direct relationships between numbers and perception somewhat impossible. Hence, human responses are fundamental in understanding the perceptual relevance of physical parameters, and in consequence reverberation properties *per se*. Having human subjects as the measuring instruments introduce interpersonal differences, based on taste, beliefs, experience, value, and need [162]. In addition, there is no common vocabulary for acoustic stimuli, and we mostly borrow words from other senses (i.e. warmth, clear, muddy) [163]. Even the most cited terminology of perceptual attributes in acoustics [36], has been described by the author as a 'non-accurate depiction'. This ambivalent interpretation of data is very limiting [18] and generalisations are problematic even when identical contextual factors are used. Fortunately these limitations have been known in other industries (e.g. food, wine) and tools like *Sensory Evaluation* [164] make it possible to extract objective information often hidden behind people's global judgments (i.e. preference).

Summary of Experimental Techniques in the Perceptual Domain

The majority of published studies in the perceptual domain aim to provide a model, where subjects make judgments about the acoustics of certain rooms. Over the last decades several techniques have been employed, which can be roughly categorised in three major groups: (1) *in-situ* evaluation (2) Auralisation / Laboratory settings (3) Combination Multichannel recording/reproduction techniques. Conducting *in-situ* evaluations of live performances in concert halls seems the most ecologically valid scenario, as its purpose and provided settings match the reality. However, it includes many uncontrolled parameters, as halls differ in a myriad of ways (i.e. shape, structure, materials), orchestra performances are rarely repeatable (i.e. dynamics, tempo), and the required test-to-test period expands to days, introducing cognitive issues (i.e. memory, mood, expectation) [165] making it difficult to pinpoint direct influences. Moreover, the standardised objective metrics, which are effectively the independent variables of such evaluation cannot provide the whole picture but averages for certain parameters of the sound field. Conducting laboratory-based evaluations overcomes numerous shortcomings of *in-situ* evaluations, as it includes standardised source, signals, listening environment, and direct comparison of different halls. However, these methods often miss realism due to lack of visual input and imperfect reproduction while subjects have certain demand characteristics (see [51]), i.e. an orchestra in a room should sound like subjects are expecting it to. It is apparent that perceptual attributes elicited in these studies relate to technical inadequacies, i.e. noise, distortion, spatial distribution. A new hybrid concept of a 'loudspeaker orchestra' presented in a series of papers by Tapio Lokki and his group (see [166]), seems to overcome several shortcomings of both techniques while sourcing the merits of both in-situ and controlled laboratory settings. This approach could be applied in a scalable system for smaller rooms.

4 Further Work

The aim of the study was to formulate our initial investigation on the perceptual attributes of reverberation, by using existing relationships as a key framework and benchmark, towards further experiments in the perceptual control of reverberant sound fields. It is apparent that perception is highly important in understanding reverberation, but dealing with perception includes various biases, as well as a trade-off between controlled settings, and ecological validity.

Sensory Analysis methodologies found in food and wine industry [164, 167] have been successfully applied to acoustical research over the last decade [4, 7, 16, 33, 167–173] and they seem to provide more accurate information about perception of acoustical features, avoiding linguistic, subjective, and biased responses as discussed above, in a well-structured and scientific framework (see [7, 170]). The advantage of these techniques in investigating such complex phenomena lies in their ability to extract information normally hidden behind hedonic and affective judgments. It seems therefore a well-suited set of techniques for investigating human's responses to different reverberant fields in small rooms. Considerations in further work should include the standardisation of the sound source (i.e. a typical directional speaker), lifelike, dynamic and time-varying signals, systematic alteration of certain room parameters, as well as to provide as realistic scenarios as possible (i.e. providing visual input, including head-tracking). Moreover, the existence of two categories of listeners' preference in reverberant settings [4, 38, 43, 44] should also be considered; individual vocabulary techniques maybe used [76, 174]. Last but not least, the requirement of Sensory Analysis of simultaneous comparison of ' tasting samples' – i.e. this paper's rooms with different reverberant fields – suggests that these settings must be recorded and reproduced in the laboratory. *In-situ* evaluation will not reveal reliable results due to limited auditory memory, and other reasons already discussed above. Following these findings, a systematic experiment will be further performed aiming to verify perceptual attributes of reverberation and their thresholds, based on realistic, yet controlled settings.

5 Conclusions - General Remarks

This paper summarises the published literature on reverberation following a top-down approach, from high level conceptual attributes to lower level metrics, physical description of reverberant fields. This investigation included a representative set of relevant works in concert hall acoustics, in-situ and in laboratory settings, as well as psychoacoustics, acoustic measurement and spatial audio research. As the published literature is plentiful, we examined and analysed the most important results that fall in the perceptual domain of the filter model. Here we presented a summary of the most salient perceptual attributes already identified in the literature as well as any possible direct relationships to physical, and affective domains. The identified perceptual aspects presented, aim to provide not necessarily an exhaustive list, but a characteristic ranked sample of perceptual attributes and senses related to reverberation, as an attempt to provide a key framework for further research. In the interest of this study it has been apparent that several major perceptual attributes seem to relate highly to properties of the auditory system such as Precedence Effect, masking, spectral and temporal binaural processing, as well as other cues (i.e. vision, expectation, episodic memory). Such mechanisms should be considered by objective metrics, and thresholds (JNDs) should be identified for the major perceptual and objective aspects of reverberation, towards efficient and robust characterisation of reverberant sound fields.

Acknowledgment

Acknowledgments: The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ITN-GA-2012-316969 and it was supported by a Postdoctoral Fellowship of the Research Foundation Flanders (FWO-Vlaanderen).

References

[1] ISO 3382-1:2009, Acoustics—Measurement of Room Acoustic Parameters—Part 1: Performance Spaces (International Standards Organization, Geneva, Switzerland, 2009). 3, 2009.

[2] ISO 3382-2:2008, Acoustics—Measurement of Room Acoustic Parameters—Part 2: Reverberation Time in ordinary Rooms (International Standards Organization, Geneva, Switzerland, 2009). 2008.

[3] L. Beranek. Concert hall acoustics—2008. *Journal of the Audio Engineering Society*, 56(7), 2008.

[4] T. Lokki, J. Patynen, A. Kuusinen, and S. Tervo. Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *Journal of the Acoustical Society of America*, 132(5):3148–3161, 2012.

[5] J. Bradley. Review of objective room acoustics measures and future needs. *Applied Acoustics*, 72(10):713–720, Oct. 2011.

[6] T. H. Pedersen and C. Fog. Optimisation of Perceived Product Quality. In *Euronoise II*, pages 633–638, 1998.

[7] S. Bech and N. Zacharov. *Perceptual audio evaluation-Theory, method and application.* Wiley, 2007.

[8] S. Bech, R. Hamberg, M. Nijenhuis, K. Teunissen, H. Looren de Jong, P. Houben, and S. K. Pramanik. Rapid perceptual image description (RaPID) method. In B. E. Rogowitz and J. P. Allebach, editors, *Electronic Imaging: Science & Technology*, pages 317–328. International Society for Optics and Photonics, Apr. 1996.

[9] H. Kuttruff. *Room acoustics.* Spon Press, fourth edition, 2009.

[10] L. L. Beranek and T. Mellow. *Acoustics: Sound Fields and Transducers.* 2012.

[11] T. D. Rossing, P. A. Wheeler, and F. R. Moore. *The science of sound.* Addison Wesley, 2002.

[12] A. Gade. Acoustics in halls for speech and music. *Springer handbook of acoustics*, 2007.

[13] A. H. Marshall and M. Barron. Spatial responsiveness in concert halls and the origins of spatial impression. *Applied Acoustics*, 62(2):91–108, Feb. 2001.

[14] L. Cremer, H. Müller, and T. Schaultz. *Principles and applications of room acoustics.* Applied Science, London, 1982.

[15] L. Kirkegaard and T. Gulsrud. In Search of a New Paradigm: How Do Our Parameters and Measurement Techniques Constrain Approaches to Concert Hall Design? *Acoustics Today*, 7(1):7, Mar. 2011.

[16] A. Kuusinen, J. Pätynen, S. Tervo, and T. Lokki. Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls. *The Journal of the Acoustical Society of America*, 135(1):239–250, Jan. 2014.

[17] J. v. D. Schuitman. Auditory Modelling for Assesing Room Acoustics. PhD Thesis. University of Delft. 2011.

[18] T. Lokki. Throw away that standard and Listen: your two ears work better. In *International Symposium on Room Acoustics*, Toronto, Canada, 2013.

[19] D. de Vries, E. M. Hulsebos, and J. Baan. Spatial fluctuations in measures for spaciousness. *The Journal of the Acoustical Society of America*, 110(2):947, 2001.

[20] J. Bradley and R. Halliwell. Accuracy and reproducibility of auditorium acoustics measures. 1988.

[21] J. Bradley. Comparison of concert hall measurements of spatial impression. *The Journal of the Acoustical Society of America*, 96(2):3525–3535, 1994.

[22] X. Pelorson, J. Vian, and J. Polack. On the variability of room acoustical parameters: reproducibility and statistical validity. *Applied Acoustics*, 1992.

[23] T. Hidaka, L. L. Beranek, and T. Okano. Interaural cross-correlation (IACC), lateral fraction (LF), and low- and high-frequency sound levels (G) as measures of acoustical quality in concert halls. *The Journal of the Acoustical Society of America*, 97(5):3319, May 1995.

[24] D. de Vries and J. Baan. Fluctuation of room acoustical parameters on small spatial intervals. *The Journal of the Acoustical Society of America*, 105(2):1367, Feb. 1999.

[25] K. Sekiguchi and T. Hanyu. Study on acoustic index variations due to small changes in the observation point. *Proceedings of the 15th ICA Seattle*, 1998.

[26] W. C. Sabine. *The Collected Papers on Acoustics*, volume 33. University of California, Los Angeles, 1922.

[27] M. Barron. Objective assessment of concert hall acoustics using Temporal Energy Analysis. *Applied Acoustics*, 74(7):936–944, July 2013.

[28] L. Cremer. *Die wissenschaftlichen Grundlagen der Raumakustik: Band I, Geometrische Raumakustik*. Hirzel-Velgard, Stuttgard, 1948.

[29] R. Thiele. Directional Distribution of Reflections in Rooms (In German). *Acustica*, 3(1):291–302, 1953.

[30] A. Marshall. A note on the importance of room cross-section in concert halls. *Journal of Sound and Vibration*, 1967.

[31] M. Barron. The subjective effects of first reflections in concert halls—the need for lateral reflections. *Journal of sound and vibration*, 15(4):475–494, 1971.

[32] P. Damaske and Y. Ando. Interaural crosscorrelation for multichannel loudspeaker reproduction. *Acta Acustica united with Acustica*, 1972.

[33] J. Patynen, S. Tervo, P. W. Robinson, and T. Lokki. Concert halls with strong lateral reflections enhance musical dynamics. *Proceedings of the National Academy of Sciences*, 2014(15):1319976111–, Mar. 2014.

[34] C. Guastavino and B. F. G. Katz. Perceptual evaluation of multi-dimensional spatial audio reproduction. *The Journal of the Acoustical Society of America*, 116(2):1105, 2004.

[35] D. Dubois. Categories as acts of meaning: The case of categories in olfaction and audition. *Cognitive Science Quarterly*, 1(4):35–38, 2000.

[36] L. Beranek. *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. Springer, New York, 2004.

[37] R. Hawkes and H. Douglas. Subjective acoustic experience in concert auditoria. *Acta Acustica united with Acustica*, 1971.

[38] M. Barron. Subjective study of British symphony concert halls. *Acta Acustica united with Acustica*, 66(1), 1988.

[39] E. Kahle. *Validation dåun mod'ele objectif de la perception de la qualit le acoustique dans un ensemble de salles de concerts et dåop Ierass*. PhD thesis, 1995.

[40] A. Sotiropoulou. Concert hall acoustic evaluations by ordinary concert-goers: I, Multi-dimensional description of evaluations. *Acta Acustica united with Acustica*, 81(1):10–19, 1995.

[41] A. Giménez, R. M. Cibrián, and S. Cerdá. Subjective Assessment of Concert Halls: a Common Vocabulary for Music Lovers and Acousticians. *Archives of Acoustics*, 37(3):331–340, Jan. 2012.

[42] R. Kurer, G. Plenge, and H. Wilkens. Correct Spatial Sound Perception Rendered by a Special 2-Channel Recording Method. In *Audio Engineering Society Convention 37*, Berlin, Germany, Oct. 1969. Audio Engineering Society.

[43] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse. Comparative Study of European Concert halls: correlation of Subjective preference with Geometric and Acoustic Parameters. *Journal of the Acoustical Society of America*, 56:1195–1201, 1974.

[44] G. Soulodre and J. Bradley. Subjective evaluation of new room acoustic measures. *The Journal of the Acoustical Society of America*, 98(1):294–301, 1995.

[45] O. Warusfel, C. Lavandier, and J. Jullien. Perception of coloration and spatial effects in room acoustics. *. . . of the 13th International Congress on Acoustics (. . .* , 1989.

[46] J. Bradley and G. Soulodre. Objective Measures of Listener Envelopment. *The Journal of the Acoustical Society of America*, 98(5):2590–2597, 1995.

[47] Y. Choi and F. Fricke. A comparison of subjective assessments of recorded music and computer simulated auralizations in two auditoria. *Acta acustica united with acustica*, 2006.

[48] J. Pätynen and T. Lokki. Evaluation of concert hall auralization with virtual symphony orchestra. *Building Acoustics*, 18(3-4):349–366, 2011.

[49] F. Rumsey. Subjective assessment of the spatial attributes of reproduced sound. In *15th International Conference: Audio Engineering Society*, 1998.

[50] F. Rumsey and J. Berg. Verification and correlation of attributes used for describing the spatial quality of reproduced sound. In *The 19th International Conference: Surround Sound*, 2001.

[51] F. Rumsey. Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society*, pages 651–666, 2002.

[52] J. Berg and F. Rumsey. Spatial attribute identification and scaling by repertory grid technique and other methods. *16th International Conference: Spatial Audio*, 1999.

[53] J. Berg and F. Rumsey. Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction. In *Proceedings of the 109th Convention of the Audio Engineering Society*, Pitea, 2000.

[54] J. Berg and F. Rumsey. Validity of Selected Spatial Attributes in the Evaluation of 5-channel Microphone techniques. In *The 112th Convention of the Audio Engineering Society*, Munich, 2002.

[55] N. Zacharov and K. Koivuniemi. Unravelling the perception of spatial sound reproduction: Language development, verbal protocol analysis and listener training. In *Audio Engineering Society Convention 111*, Tampere, FInland, 2001.

[56] E. Kahle. Room acoustical quality of concert halls : perceptual factors and acoustic criteria – return from experience. In *International Symposium on Room Acoustics*, 2013.

[57] B. Blesser and L. Salter. *Spaces Speak Are you Listening.* MIT Press, 2007.

[58] P. Huang, J. Abel, H. Terasawa, and J. Berger. Reverberation Echo Density Psychoacoustics. *Proceedings of the 125th Convention of the Audio Engineering Society*, 2008.

[59] V. P. Sivonen. Directional loudness and binaural summation for wideband and reverberant sounds. *The Journal of the Acoustical Society of America*, 121(5):2852–61, May 2007.

[60] T. Lokki, J. Patymen, S. Tervo, S. Siltanen, and L. Savioja. Engaging concert hall acoustics is made up of temporal envelope preserving reflections. *Journal of the Acoustical Society of America - Express Letters*, 129(6):EL223–8, June 2011.

[61] A. W. Bronkhorst and T. Houtgast. Auditory distance perception in rooms. *Nature*, 397(6719):517–520, 1999.

[62] S. Hameed, J. Pakarinen, K. Valde, and V. Pulkki. Psychoacoustic Cues in Room Size Perception. *116th Convention of the Audio Engineering Society*, pages 1–7, 2004.

[63] A. Tajadura-Jiménez, P. Larsson, A. Väljamäe, D. Västfjäll, and M. Kleiner. When room size matters: acoustic influences on emotional responses to sounds. *Emotion (Washington, D.C.)*, 10(3):416–22, June 2010.

[64] P. N. Juslin and D. Västfjäll. Emotional responses to music: the need to consider underlying mechanisms. *The Behavioral and brain sciences*, 31(5):559–75; discussion 575–621, Oct. 2008.

[65] M. Bitner. Servicescapes: The impact of physical surroundings on customers and employees. *Journal of marketing*, 1992.

[66] EBU Rec. 562-3, "Subjective Assessment of Sound Quality," European Broadcasting Union, Geneva, Switzerland (1990). Technical report.

[67] M. Yadav, D. Cabrera, L. Miranda, W. L. Martens, D. Lee, and R. Collins. Investigating Auditory Room Size Perception with Autophonic Stimuli. In *The 135th Convention of the Audio Engineering Society*, volume 39, pages 101–105, New York, USA, 2013. Audio Engineering Society.

[68] T. Lokki, J. Patynen, A. Kuusinen, H. Vertanen, and S. Tervo. Concert hall acoustics assessment with individually elicited attributes. *The Journal of the Acoustical Society of America*, 130(2):835–49, Aug. 2011.

[69] J. Sandvad. Auditory perception of reverberant surroundings. *The Journal of the Acoustical Society of America*, 105(2):1193, Feb. 1999.

[70] E. Calcagno, E. Abregu, M. Eguía, and R. Vergara. The role of vision in auditory distance perception. *Perception*, 2012.

[71] C. Pop and D. Cabrera. Auditory room size perception for real rooms. *Proceedings of ACOUSTICS, Australia*, (November), 2005.

[72] D. Mershon and L. King. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 1975.

[73] C. Mendonca, J. Lamas, T. Barker, G. Campos, P. Dias, V. Pulkki, C. Silva, and J. a. Santos. Reflection orders and auditory distance. 19:050041–050041, 2013.

[74] M. Barron. Late lateral energy fractions and the envelopment question in concert halls. *Applied Acoustics*, 62(2):185–202, Feb. 2001.

[75] T. Letowski. Sound Quality Assessment: Cardinal Concepts. *87th Convention of the Audio Engineering Society*, 1989.

[76] G. Lorho. Evaluation of spatial enhancement system for Stereo Headphone Reproduction by Preference and Attribute Rating. In *Proceedings of the 118th Convention of the Audio Engineering Society*, 2005.

[77] A. Gabrielsson and H. Sjögren. Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, 65(4):1019–33, Apr. 1979.

[78] M. Slater, V. Linakis, M. Usoh, and R. Kooper. Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. *ACM virtual reality software*, 1996.

[79] M. Slater, M. Usoh, and A. Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction*, 1995.

[80] T. Hidaka and L. Beranek. Objective and subjective evaluations of twenty-three opera houses in Europe, Japan, and the Americas. *The Journal of the Acoustical Society of America*, 107(1):368–83, Jan. 2000.

[81] R. C. Heyser. Acoustical Measurements by Time Delay Spectrometry. *Journal of the Audio Engineering Society*, 15(4):370–382, Oct. 1967.

[82] D. Davis. The Role of the Initial Time Delay Gap in the Acoustic Design of Control Rooms for Recording and Reinforcement Systems. *Audio Engineering Society Convention 64*, 1979.

[83] D. Davis and C. Davis. The LEDE-Concept for the Control of Acoustic and Psychoacoustic Parameters in Recording Control Rooms. *Journal of the Audio Engineering Society*, 1980.

[84] J. Li and R. Wu. An efficient algorithm for time delay estimation. *IEEE Transactions on Signal Processing*, 46(8):2231–2235, 1998.

[85] W. W. Renbiao Wu and Q. Jia. FFT-Based efficient algorithms for time-delay estimation. In S. Salih, editor, *Fourier Transform - Signal Processing*, pages 55–80. InTech, Apr. 2012.

[86] J. M. Potter. On the Binaural Modelling of Spaciousness in Room Acoustics. *Phd Thesis*, 1993.

[87] D. Griesinger. Objective Measures of Spaciousness and Envelopment. *Audio Engineering Society Conference 16th International Conference Spatial Sound Reproduction*, pages 1–15, 1999.

[88] M. Barron. Using the standard on objective measures for concert auditoria, ISO 3382, to give reliable results. *Acoustical Science and Technology*, 26(2):162–169, 2005.

[89] S. Kim, D. Ko, A. Nagendra, and W. Woszczyk. Subjective Evaluation of Multichannel Sound with Surround-Height Channels. In *Audio Engineering Society Convention 135*, pages 1–7. Audio Engineering Society, 2013.

[90] M. Morimoto and Z. Maekawa. Auditory spaciousness and envelopment. *Proc. 13th Int. Congr. on Acoustics*, 1989.

[91] M. Barron and a.H. Marshall. Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. *Journal of Sound and Vibration*, 77(2):211–232, July 1981.

[92] J. Blauert, W. Lindemann, R.-U. Bochum, and F. Republic. Auditory spaciousness: Some further psychoacoustic analyses. *The Journal of the Acoustical Society of ...*, pages 533–542, 1986.

[93] W. Kuhl. Räumlichkeit als komponente des raumeindrucks. *Acta Acustica united with Acustica*, 1978.

[94] C. Kim, R. Mason, and T. Brookes. Head movements made by listeners in experimental and real-life listening activities. *Journal of the Audio Engineering Society*, pages 425–438, 2013.

[95] S. Choisel and F. Wickelmaier. Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound. *Journal of the Audio Engineering Society*, pages 815–826, 2006.

[96] H. Wittek. Perceptual differences between wavefield synthesis and stereophony. 2007.

[97] R. Conetta, F. Rumsey, and S. Zielinski. QESTRAL (Part 2): Calibrating the QESTRAL model using listening test data. In *Proceedings of the 125 Convention of the Audio Engineering Society*, San Fransisco, 2008.

[98] F. Rumsey, S. Zielinski, and P. Jackson. QESTRAL (Part 1): Quality evaluation of spatial transmission and reproduction using an artificial listener. In *Proceedings of the 125 Convention of the Audio Engineering Society*, San Fransisco, 2008.

[99] M. Morimoto, K. Iida, and K. Sakagami. The role of reflections from behind the listener in spatial Impression. *Applied Acoustics*, 62(1):109–124, 2001.

[100] IEC. 268-13, Recommendation for Listening Room. Technical report, Geneva, Switzerland, 1985.

[101] T. Okano, L. L. Beranek, and T. Hidaka. Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls. *The Journal of the Acoustical Society of America*, 104(1):255–65, July 1998.

[102] J. Blauert. *Spatial hearing: the psychophysics of human sound localization.* MIT Press, Cambridge, revised ed edition, 1997.

[103] J. Blauert and W. Lindemann. Spatial Mapping of Intracranial Auditory Events for various Degrees of Interaural Coherence. *Journal of the Acoustical Society of America*, 79(3):806–813, 1986.

[104] W. Lindemann. Extension of a binaural crosscorrelation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America*, 1986.

[105] W. Lindemann. Extension of a Binaural Cross-Correlation Model by Means of Contralateral Inhibition. II. The Law of the First Wave Front. *Journal of Acoustical Society of America*, 1986.

[106] R. Mason. *Elicitation and measurement of auditory spatial attributes in reproduced sound. PhD Thesis.* PhD thesis, University of Surrey, 2002.

[107] J. Becker. Spectral and temporal contribution of different signals to ASW analysed with binaural hearing models. *Proceedings of the Forum Acusticum*, 2002.

[108] W. Keet. The influence of early lateral reflections on the spatial impression. *Actes du 6e Congrès international d'Acoustique, Tokyo*, 1968.

[109] F. Bilsen. Psychophysical observations on spaciousness. *10th International Congress of Acoustics, Sydney*, 1980.

[110] F. H. Bilsen and van den Berg. Spaciousness and its(psycho-) acoustical parameters. In *Proceedings of the 12 International Congress on Acoustics*, Toronto, Canada, 1986.

[111] D. Griesinger. The Psychoacoustics of Apparent Source Width , Spaciousness and Envelopment in Performance Spaces. *Acta Acustica united with Acustica*, 83(Feb):721–731, 1997.

[112] W. Kuhl. In der Raumakustik benutzte hürakustische Termini. *Acustica*, 39:57–58, 1977.

[113] W. Reichardt and U. Lehmann. Raumeindruck als Oberbegriff von Räumlichkeit und Halligkeit, Erläuterungen des Raumeindrucksmaßes R. *Acta Acustica united with Acustica*, 40(5), 1978.

[114] A. M. Sarroff and J. P. Bello. Toward a Computational Model of Perceived Spaciousness in Recorded Music . 59(7), 2011.

[115] S. Bhariktar and C. Kyriakakis. *Immersive Audio Signal Processing.* Springer, 2006.

[116] J. S. Bradley, R. D. Reich, and S. G. Norcross. On the combined effects of early- and late-arriving sound on spatial impression in concert halls. *The Journal of the Acoustical Society of America*, 108(2):651, Aug. 2000.

[117] H. Furuya, K. Fujimoto, A. Wakuda, and Y. Nakano. The influence of total and directional energy of late sound on listener envelopment. *Acoustical science and Technology*, 2005.

[118] K. Ueda, T. Tanaka, and M. Morimoto. Estimation of auditory source width (ASW): II. ASW for two adjacent 1/3 octave band noises with different band levels. *Journal of Acoustical Society Society of Japan*, 18(3):121–128, 1997.

[119] K. Ueda and M. Morimoto. Estimation of auditory source width (ASW): I . ASW for two adjacent 1 / 3 octave band noises with equal band level. *Journal of the Acoustical Society of Japan (E)*, 16(2):77–83, 1995.

[120] S. Choisel and F. Wickelmaier. Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. *The Journal of the Acoustical Society of America*, 121(1):388, 2007.

[121] F. E. Toole. Subjective measurements of loudspeaker sound quality and listener performance. *Journal of the Audio Engineering Society*, 33(1), 1985.

[122] I. Recommendation. 60268, Sound System Equipment— Part 13: Listening Tests on Loudspeakers, International Electrotechnical Commission, Geneva, Switzerland (1985). *International Electrotechnical Commission*, 1985,1997.

[123] F. E. Toole. *Sound Reproduction: Loudspeakers and Rooms.* Focal Press, Bostom, 2008.

[124] T. Lokki and J. Pätynen. Lateral reflections are favorable in concert halls due to binaural loudness. *The Journal of the Acoustical Society of America*, 130(5):EL345–51, Nov. 2011.

[125] H. Lee. Apparent Source Width and Listener Envelopment in Relation to Source-Listener Distance. *52nd International Conference: Sound Field Control, Audio Engineering Society*, pages 1–6, 2013.

[126] T. Hidaka, L. L. Beranek, and T. Okano. Some considerations of interaural cross correlation and lateral fraction as measures of spaciousness in concert halls. Chapter 32 - Music and Concert Hall Acoustics (Ando and Noson), England UK: Academic Press. 1997.

[127] S. E. Olive and F. E. Toole. The Detection of Reflections in Typical Rooms. *Journal of the Audio Engineering Society*, 37(7/8):539–553, July 1989.

[128] J. van Dorp Schuitman, D. de Vries, and A. Lindau. Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model. *The Journal of the Acoustical Society of America*, 133(3):1572–85, Mar. 2013.

[129] W. Hess. *Time-variant Binaural Activity Characteristics as Indicator of Auditory Spatial Attributes. PhD Thesis.* Ruhr-Universitat Bochum, Bochum, 2006.

[130] D. Robinson and L. Jeffress. Effect of varying the interaural noise correlation on the detectability of tonal signals. *The Journal of the Acoustical Society of America*, 1963.

[131] N. I. Durlach. Interaural correlation discrimination: II. Relation to binaural unmasking. *The Journal of the Acoustical Society of America*, 79(5):1548, May 1986.

[132] M. Morimoto and K. Iida. A practical evaluation method of Auditory Source Width in concert halls. *Journal of the Acoustical Society of Japan (E)*, 16(2):59–69, 1995.

[133] F. L. Wightman. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648, Mar. 1992.

[134] D. Griesinger. Room impression, reverberance, and warmth in rooms and halls. *Audio Engineering Society Convention 93*, 1992.

[135] M. Morimoto, S. Sugiura, and K. Iida. Relation between auditory source width in various sound fields and degree of interaural cross-correlation: Confirmation by constant method. *Applied Acoustics*, 1994.

[136] M. Morimoto. The relation between auditory source width and the law of the first wave front. *Proc. Inst. Acoust.*, 1992.

[137] M. Morimoto, K. Iida, and Y. Furue. Relation between auditory source width in various sound fields and degree of interaural cross-correlation. *Applied Acoustics*, 38(2-4):291–301, Jan. 1993.

[138] S. Klockgether, J. van Dorp Schuitman, and S. Van de Par. Perceptual Limits for detecting interaural-cue manipulations measured in reverberant settings. *Proceedings of Meetings on Acoustics*, 19:015004–015004, 2013.

[139] A. Walther and C. Faller. Interaural correlation discrimination from diffuse field reference correlations. *The Journal of the Acoustical Society of America*, 133(3):1496–502, Mar. 2013.

[140] W. Martens. The impact of decorrelated low-frequency reproduction on auditory spatial imagery: Are two subwoofers better than one? *Proc. AES 16th International Conference: Spatial Sound Reproduction*, (67-77), 1999.

[141] S. Linkwitz. Room Reflections Misunderstood. *Audio Engineering Society Convention 123*, 2007.

[142] M. B. Gardner. Distance Estimation of 0◦ or Apparent 0◦ -Oriented Speech Signals in Anechoic Space. *The Journal of the Acoustical Society of America*, 45(1):47, July 1969.

[143] T. Lokki, J. Pätynen, and N. Zacharov. Concert Hall Acoustics Assessment with Sensory Evaluation - Tools and Practices. *Prooceedings of the Institute of Acoustics*, 33(2):35–39, 2011.

[144] B. Shinn-Cunningham. Learning reverberation: Considerations for spatial auditory displays. *International Conference on Auditory Display*, 2000.

[145] D. Mershon, W. Ballenger, and A. Little. Effects of room reflectance and background noise on perceived auditory distance. *Perception*, (18:3):403–408, 1989.

[146] A. Bronkhorst. Modeling auditory distance perception in rooms. *Proc. EAA Forum Acusticum Sevilla*, 2002.

[147] S. H. Nielsen. Auditory Distance Perception in Different Rooms. *Journal of the Audio Engineering Society*, 41(10):755–770, Oct. 1993.

[148] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832, 2002.

[149] A. J. Kolarik, S. Cirstea, and S. Pardhan. Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 224(4):623–33, Feb. 2013.

[150] P. D. Coleman. Failure to localize the source distance of an unfamiliar sound. *The Journal of the Acoustical Society of America*, 34:345–346, 1962.

[151] P. Coleman. An analysis of cues to auditory depth perception in free space. *Psychological Bulletin*, 1963.

[152] P. Zahorik. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91(February 2003):409–420, 2005.

[153] P. Zahorik and F. Wightman. Loudness constancy with varying sound source distance. *Nature neuroscience*, 2001.

[154] M. Barron. When is a concert hall too quiet. *Proceedings of the 19th International Congress on Acoustics*, 2007.

[155] J. Philbeck and D. Mershon. Knowledge about typical source output influences perceived auditory distance. *The Journal of the Acoustical Society of America*, 111:1980–1983, 2002.

[156] B. C. J. Moore. *An Introduction to the Psychology of Hearing.* Emerald, sixth edition, 2013.

[157] O. J. Pedersen, S. r. Bech, K. Rasmussen, T. U. o. D. A. Laboratory, and DTU. *Final Report on Activities of Eureka Project ARCHIMEDES.* DTU press, 1995.

[158] J. Blauert and U. Jekosch. Auditory quality of concert halls-the problem of references. *Proc. 19-th Int. Congr. Acoust., ICA*, 2007.

[159] I. Rock, C. Linnett, P. Grant, and A. Mack. Perception without attention: Results of a new method. *Cognitive Psychology*, 1992.

[160] D. D. L. Valente, J. Braasch, and N. York. Subjective scaling of spatial room acoustic parameters influenced by visual environmental cues. *The Journal of the Acoustical Society of America*, 128(4):1952–64, Oct. 2010.

[161] D. L. Valente, J. Braasch, and S. a. Myrbeck. Comparing perceived auditory width to the visual image of a performing ensemble in contrasting bi-modal environments. *The Journal of the Acoustical Society of America*, 131(1):205–17, Jan. 2012.

[162] V. Hulusic, C. Harvey, K. Debattista, N. Tsingos, S. Walker, D. Howard, and A. Chalmers. Acoustic Rendering and Auditory-Visual Cross-Modal Perception and Interaction. *Computer Graphics Forum*, 31(1):102–131, Feb. 2012.

[163] L. E. Marks. Bright sneezes and dark coughs, loud sunlight and soft moonlight. *Journal of experimental psychology. Human perception and performance*, 8(2):177–93, Apr. 1982.

[164] H. T. Lawless and H. Heymann. *Sensory Evaluation of Food: Principles and Practices.* 1999.

[165] M. Sams, R. Hari, J. Rif, and J. Knuutila. The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence. *Journal of Cognitive Neuroscience*, 5(3):363–368, Dec. 1993.

[166] J. Pätynen. *A virtual symphony orchestra for studies of concert hall acoustics.* PhD thesis, Aalto University School of Science, 2011.

[167] C. Simonsen and S. Legarth. A procedure for sound quality evaluation of hearing aids. *Hearing Review, Dec*, 17(13):32–37, 2010.

[168] G. Lorho. Perceived Quality Evaluation - An application to sound reproduction over headphones. *Phd Thesis*, 2010.

[169] T. Lokki, J. Pätynen, and N. Zacharov. Concert Hall Acoustics Assesment with Sensory Evaluation - Tools and Practices. In *Proceedings of the Institute of Acoustics*, volume 33, 2011.

[170] T. Lokki. Tasting music like wine: Sensory evaluation of concert halls. *Physics Today*, 67(1):27–32, Jan. 2014.

[171] N. Zacharov, J. Ramsgaard, G. Le Ray, and C. V. Jorgensen. The multidimensional characterization of active noise cancelation headphone perception. In *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 130–135. IEEE, June 2010.

[172] J. Berg and F. Rumsey. Identification of quality attributes of spatial audio by repertory grid technique. *Journal of the Audio Engineering Society*, pages 365–379, 2006.

[173] T. Neher, T. Brookes, and F. Rumsey. A Hybrid Technique for Validating Unidimensionality of Perceived Variation in a Spatial Auditory Stimulus Set. *Journal of the Audio Engineering Society*, 54(4):259–275, Apr. 2006.

[174] A. Williams and G. Arnold. A comparison of the aromas of six coffees characterised by conventional profiling, free choice profiling and similarity scaling methods. *Journal of the Science of Food and Agriculture*, 1985.

Paper B

A Rapid Sensory Analysis Method for Perceptual Assessment of Automotive Audio

Neofytos Kaplanis, Søren Bech, Sakari Tervo, Jukka Pätynen,
Tapio Lokki, Toon van Waterschoot, and Søren Holdt Jensen.

The paper was published in the
Journal of Audio Engineering Society Vol. 65(1/2), pp. – , 2017.

The layout has been revised.

Abstract

Today's car audio systems encompass some of the latest sound technologies available, capable of delivering unique and novel aural experiences. It remains unclear whether current perceptual assessment protocols follow this trend and their ability to fully capture the human sensations evoked by such systems is questioned. This paper reviews the applicability of existing assessment protocols in today's systems and draws upon the identified limitations. It further reports the design and implementation of a new method to perceptually investigate the properties of automotive audio. The method uses Spatial Decomposition Method for acquiring, analyzing, and reproducing the sound field in a laboratory over loudspeakers, allowing instant comparisons of automotive audio systems. A Rapid Sensory Analysis protocol, the Flash Profile, is employed for evaluating the perceptual experience using individually elicited attributes, in a time-efficient manner. A pilot experiment is presented, where experts, experienced, and naive assessors followed the proposed procedure and evaluated three sound fields. The current findings suggest that the method allows the assessment of both spatial and timbral properties of automotive audio. This may form a scientific framework for characterizing the acoustical qualities within the automotive environment and stipulate research paths to better understand these sound fields.

1 Introduction

Over the last decades the automotive industry has been focusing on identifying and improving the major factors that influence the sensory experience within the vehicles. As a consequence, the study of sound quality in automotive audio systems has been brought into the limelight [1]. Although sound quality research has established and standardized a plethora of assessment procedures [2, 3], protocols for automotive audio are yet to be defined. Here, the published literature on the past and current practices is reviewed, aiming to stipulate new approaches and encourage a general framework for automotive audio assessment.

The highly complex and acoustically hostile environment of a car cabin [4–6] hinders the effectiveness of standard objective metrics [7], lacking robustness, repeatability, and perceptual relevance [8, 9]. This has naturally led to the use of the human auditory system as a major instrument in developing and evaluating car audio. Aiming towards a high quality aural experience, car audio manufacturers normally employ perceptual assessment protocols to characterize and optimize these sound fields [10].

In the lack of standardized evaluation procedures [11], automotive audio assessment protocols adopted paradigms from sound quality research in rooms.

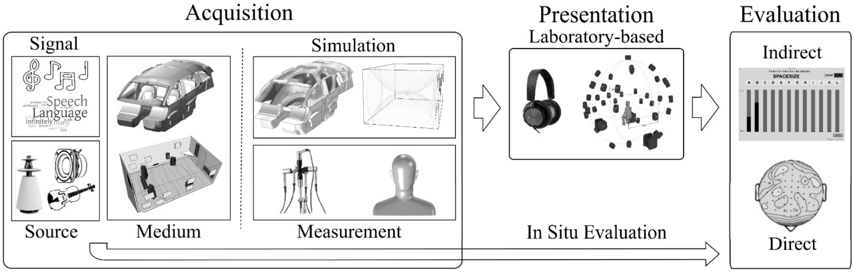


Fig. B.1: Basic principles of perceptual audio assessment.

As a consequence, the majority of the studies have been focusing on comparative evaluation of system properties. For example, towards the electroacoustic properties of the transducers, signal processing algorithms, and equalization settings [11–17], whilst attaining a perceptual experience similar to a conventional listening space, such as a mixing studio or a listening room [13, 18].

Yet, the car cabin is far from a common listening environment [13, 14, 19]. Compared to a typical listening room [2], the car cabin is characterized by a small volume, highly reflective surfaces that contrast with the inner absorptive upholstery, complex geometry, limited and often sub-optimal placement of the sound sources, as well as asymetric and dissimilar acoustical paths.

In addition, current car audio systems comprise of highly advanced loudspeaker grids [1] capable of delivering novel and environment-specific aural experiences. This trend surpassed the development of sound reproduction in typical rooms, increasing the dissimilarity of the two scenarios.

One could therefore challenge the applicability of the current perceptual assessment protocols for automotive audio evaluation [20]. New experimental frameworks should be developed, with the restrictions and specificity of the automotive environment in mind. That is, methods that would allow the identification and quantification of human perception even when the properties of the sound fields are foreign to the human ear, i.e., percepts that are difficult to estimate *a priori.* Such approaches may enhance our understanding of these fields, and stipulate perceptually relevant ways to address the subsequent degradation.

This paper proposes a research framework for perceptual assessment of automotive audio, targeting the evaluation of both the properties of the reproduction system and the acoustical properties of the car's interior (e.g. cabin). First, a brief literature review is presented in Section 2 where past and current protocols are discussed, their limitations are identified and potential improvements are proposed. A new experimental method is then presented (Sec. 3) including a pilot verification experiment (Sec. 4) where the initial results are shown. Several remarks of the method including limitations and future work are discussed (Sec. 5), followed by the concluding remarks of the study (Sec. 6).

2 Background

2.1 Perceptual Assessment in Automotive

The aim of perceptual assessment is to investigate the relationships between measurable physical quantities, and the perceived sensations. In the domain of audio these protocols comprise of three major components: the *acquisition* of the signal to be investigated, its *presentation* to human assessors, and the *evaluation* protocol where assessors' responses are collected. The principles of audio assessment protocols are summarized in Figure B.1. Typically, each experimental protocol follows the given contextual factors and limitations of the study in a way that the physical data collected during acquisition and the perceptual data collected during evaluation could be combined, and a relationship could be established.

Audio research has demonstrated a wealth of protocols for assessing audio material, which one could categorize into two major groups: the *In-situ* and the *Laboratory*-based methods. In-situ evaluations describe the protocols where the assessor evaluates the perceived sound in the intended natural settings, e.g., listening to an orchestra in a concert hall. In the laboratory-based methods, the signals are captured via measurements or simulations, and a presentation scheme is followed to recreate the sound field in a neutral environment, e.g. in a laboratory via headphones or loudspeakers.

In the following section, the literature on perceptual assessment of automotive audio will be examined, following the three basic elements of audio assessment, the acquisition, presentation, and evaluation. The analysis aims to identify the limitations of current practices and stipulate new approaches that would overcome such constraints.

2.2 Acquisition & Presentation

In-Situ Methods

Considering the complex acoustics and the multi-sensory experience of a car cabin, the in-situ evaluation is the most straightforward approach. In fact, automotive industry has been conducting in-situ evaluations since the early days of car audio. Though, it was soon realized that exposing assessors in a car cabin introduces strong biases caused by non-acoustical factors, e.g., size, price, brand, interior materials [11, 21, 22], that are likely to affect assessors' judgments. Shively [14] proposed a *blind in-situ* procedure where the non-acoustic feedback of both the interior and the exterior of the car was highly controlled.

Later, a ‘placebo’ [23] method was introduced aiming to force assessors into evaluating stimuli in random phases, under an in-situ sighted protocol.

A major shortcoming of in-situ assessment protocols, is the restricted ability of conducting instant and double-blind comparative evaluations [21]. Cecchi et al. [24] addressed this, by proposing a cabin-based apparatus that allowed the evaluation of signal processing algorithms under simultaneous and controlled comparative protocols, yet within the natural environment. However, the method cannot be extended in investigations where physical alterations are in question, e.g., cabin acoustics or loudspeaker placement.

One could argue that several car audio systems could be compared via in-situ methods, as far as the contextual factors are identical. Yet, the access to prototype cars is limited and such approach has not been followed [12, 13]. Instead, Olive [18] proposed the use of a reference listening scenario in a room, as a common comparison baseline against automotive audio systems.

Nonetheless, these protocols inherently include long test-to-test periods between the different cars or system settings. Based on the restricted auditory memory [25], it is likely that in-situ evaluations may influence the experimental results in an uncontrolled manner, for example following the assessor’s mood and expectations [20, 26]. Moreover, comparing multiple systems or cars would require repeated and sequential experiments, hence, longer experimental times.

Laboratory-based Methods

The direct response to the constraints of in-situ protocols was the development of laboratory-based methods. Laboratory-based methods aim to impose higher control on both auditory and non-auditory parameters, whilst ensuring repeatability and scientific validity. This is normally achieved by providing standardized source and signals, fixed listening environments and settings, and simulations or measurements to capture the car’s sound field.

Although utilizing acoustical simulations in automotive audio assessment is highly anticipated [27, 28], current methods are not able to adequately characterize the cabin’s acoustical field [29]. Automotive audio has therefore focused on measurement-based methods. The majority of these methods require the acquisition of the sound field in a car cabin using microphones, typically employing dummy-head recordings. The obtained signals are then used to reproduce the binaural field over headphones. An extension of the method has been also proposed, the so-called *Binaural Car Scanning* (BCS) [11–13, 21, 30]. BCS allows for natural head-movements during evaluation, by dynamically updating the appropriate measurement angle, based on the assessor’s head-orientation [31] during the evaluation. BCS overcomes the practical limitations of in-situ evaluations, and allows simultaneous comparisons between car cabins or audio systems, in a simple and repeatable framework. It therefore remains a valuable tool for car audio assessment, especially when the investigation focuses on a quantification of perceived intensity differences between systems.

The fundamental disadvantages of binaural techniques include the acoustical modification of the acquired signals due to the physical properties of the dummy-head and the inherent necessity of headphone-based playback. That is, the *Head-Related Transfer Function* (HRTF) of the dummy-head used during acquisition, may not fit the anthropometric properties of the human assessor during evaluation. This reduces the degree of *externalization* [32] of the sound source, leading to what is known as in-the-head perception of sound. This unnatural effect limits the degree of true perception of the investigated field [33]. Hegarty et al. [12], have shown that several spatial alterations in car audio systems were interpreted as timbral, even by expert assessors, questioning the validity of BCS when the spatial properties of the field are investigated.

Moreover, the physical constraints of headphone transducers limit the extent on which low frequencies could be reproduced accurately. This results in timbral and level differences between the acquired and the presented field [34]. Also, it restricts 'whole-body' vibrations, a sensation known to influence the listener preferences in binaural reproduction of rooms as well as car cabins [35, 36].

Obtaining *Impulse Responses* (IRs) with a dummy-head imposes several practical limitations. The analysis of the captured signal is restricted due to the embedded HRTFs, increasing the difficulty of applying standard analysis metrics. Moreover, the requirement for repeated measurements at different dummy-head orientations yields hundreds of measurements per source. Such acquisition may last several hours for a stereophonic setup. Even when it is desired to capture single transducer IRs, the thermal effects of the transducers, and the cabin, would make it rather difficult to perform accurate repeated measurements, suitable for BCS. Therefore, BCS methods of individually measured transducers are not followed.

2.3 Evaluation - Experimental Design

The final stage of a perceptual assessment comprises the evaluation of the stimuli. Typically it follows an experimental procedure where assessors' evoked sensations are quantified. The majority of audio evaluation protocols lies upon the indirect collection of human responses in verbal, graphical, or written form [37].

Many studies on automotive audio focused on the evaluation of the global quality of the sound field i.e. using *Basic Audio Quality* (BAQ) protocols and Preference scales [38]. Later, studies expanded on parametric evaluations, where the specific sound characteristics that influence the assessors' preference were sought, by evaluating audio material based on *perceptual attributes*. Initially these investigations adopted attributes from other domains of audio [23, 39, 40], followed by the implementation of *elicitation techniques* [41, 42] where a number perceptual attributes for car audio systems were identified [10, 12, 43].

2.4 Summary & Motivation

By examining the literature one could realize that automotive audio assessment protocols balance on a trade-off between the requirements of high *ecological validity* and direct, single-dimensional control of the contextual factors and parameters.

In car cabins, the complex sound field requires stimuli acquisition in the form of measurements, or in-situ evaluation under real conditions. When assessing car audio in-situ, a number of non-auditory features introduce many restrictions during evaluation. Similar to other streams of audio evaluation [20], following a laboratory-based approach where a captured signal is recreated in neutral conditions seems to provide flexible and well-controlled experimental designs. For automotive audio, such method would allow the assessment of several automotive audio systems, cabins, and their properties in a comparative double-blind protocol. These benefits can be considered superior to the in-situ methods, which rely on assumptions of flawless auditory memory.

Evaluating the general satisfaction of human perception is useful for benchmarking processes, yet, the specific sound characteristics that influence the subject's preference remain unknown. Using verbal descriptors, known as perceptual attributes, one could focus on singular auditory percepts of the sound field, avoiding cognitive biases relating to human's expectations, preference, and sentiments [37]. As car audio systems may deliver novel experiences to the human ear and specific to the car environment, such techniques would improve our understanding of the perceived field, by providing sensory information normally hidden behind hedonic and affective judgements.

It is therefore desirable to develop new experimental designs that would:

1. Be able to evaluate both the physical and the electrical properties of automotive audio systems, e.g. electroacoustical modifications and acoustical properties of the car's interior, in simultaneous, double-blind comparative protocol.

2. Provide means for physical quantification of such alterations.

3. Employ acquisition and presentation schemes where both the spatial and timbral characteristics of the field are preserved and provide flexibility to commonly used rendering schemes.

4. Allow the evaluation of uncommon and novel sound experience; still, meet the practical limitations of the automotive environment, i.e. fast, flexible, and efficient in both acquisition and evaluation phases.

3 Experimental Method

This section describes the implementation of a new methodology for acquiring, presenting, and evaluating automotive sound. It is the paper's intention to provide a general research protocol for the automotive audio, where the limitations identified in Sec. 2.4 are addressed. Here, several steps were taken towards a more adaptable and flexible perceptual assessment method, by incorporating new approaches in both the acquisition and presentation of the sound field, as well as the evaluation procedures.

The acquisition and presentation stages of the method are based on the *Spatial Decomposition Method* (SDM) [44, 45]. SDM is a spatial analysis and synthesis scheme, where the IRs obtained by a compact microphone array are analyzed parametrically, in terms of instantaneous pressure, time, and direction of arrival. Applying SDM in cars may introduce three major advantages over the previously discussed methods. First, the spatiotemporal analysis of the SDM data may enable a better understanding of the behavior of the cabin's sound field. The additional physical quantities and visualization capabilities of SDM [46], could be used as a physical metric when the perceived spatial properties are investigated. Second, it allows reproduction of the analyzed sound field over loudspeakers, addressing several issues of headphone-based playback. The spatial responses can be then reproduced with any rendering scheme e.g. *Vector Based Amplitude Panning* (VBAP) and *High Order Ambisonics* (HOA). Last but not least, SDM synthesis makes use of a single omnidirectional pressure microphone rather than beamforming or directional processing techniques. That is, the pressure used to synthesize the field originates from a single omnidirectional microphone, whereas the DOA calculation uses all six microphones on the array. In consequence, the reproduced sounds is not altered (e.g. colored) by the characteristics of the receiver, as commonly encountered when using directional microphones and dummy-head apparatus.

The requirements identified for the evaluation procedures for automotive audio assessment, seem to depict the need of *Descriptive Sensory Analysis* (DA) [41] techniques. DA combines the sensory characterization of the stimuli and the quantitative rating of the associated perceptual attributes within the same framework. Such methods have been successfully applied in audio, e.g., in concert halls [47, 48], spatial audio reproduction through loudspeakers [30, 42, 49] and headphones [50], hearing aids [51], and active noise cancellation [52]. DA encompasses *attribute elicitation* methods where human assessors are able to epitomize and appropriately quantify their sensations for the given set of stimuli, by defining their own perceptual attributes. Therefore, it allows the perceptual assessment of novel experiences, without the need of *a priori* quantification of the evaluation attributes. That is, the requirement of the experimenter to pre-select possible attributes to be used by assessors as scales during a parametric evaluation.

Although such techniques seem to suit the needs of automotive audio assessment, they require extensive training per product, as well as multiple sessions per assessor [41]. The time restrictions within the automotive environment limit the use of common DA techniques, even if their outcome would be ideal. Addressing this time limitation, one could employ the recently developed *Rapid Sensory Profiling* techniques [26]. In this paper, *Flash Profile* (FP) [53, 54], the most closely related rapid method to conventional DA profiling [26], is adapted for audio evaluation within the automotive environment. FP allows the listener to quickly elicit new, and non-limited attributes, which is a significant advantage compared to lengthy consensus attribute elicitation techniques [41] or fixed attribute lists (e.g. [2]) that may not reveal the full perceptual experience of the presented stimuli.

The methods and experimental procedure in the following sections serve as an example of the proposed framework. The following sections provide the details of each stage of the method, followed by a pilot experiment, for reasons of completion, as an implementation example of the proposed procedure.

3.1 Acquisition - In-situ Car Measurements

In order to obtain the acoustic characteristics of a sound reproduction system in a car cabin, in-situ recordings are required. For this study, measurements were conducted in a four-door sedan equipped with 17 band-limited transducers (5 tweeters, 7 mid-range transducers, 4 woofers, 1 subwoofer) and a multichannel automotive amplifier. The feed to the individual transducers was post-processed (i.e compensation delays, equalization) to represent a typical production car, equipped with a tuned, premium sound system.

An open spherical microphone array (G.R.A.S VI-50) comprising of two coincidental microphones on each axis, separated by 25 mm, was positioned at the driver's seat, at the average seating position [55]. The microphone probe was aligned to match the position of a dummy-head seating in the car - the center point of the head and ears' height. The distance between the microphone array and the headrest was set to 15 cm.

The IRs were measured in a way that the (electrical) input to the amplifier was measured by the electrical output of the microphones in the cabin. This type of measurements is referred to as *Vehicle Impulse Response* (VIR).

The VIRs were measured for each transducer, using a 5 s logarithmic sine-sweep method [56] at 192 kHz sampling rate using an RME UCX multichannel sound interface. The measurements were performed at 82 dB (C-weighted RMS) estimated using the forward facing microphone of the array, with system default settings. The electrical output of the measurement system was kept constant for all transducers. The car measurements were conducted in a temperature and noise regulated garage at Bang & Olufsen's premises.

3.2 Presentation

Spatial Analysis and Synthesis of VIRs

The spatial analysis and synthesis procedures followed in the current paper, are described in detail in a recent report [45], where the applicability of the SDM in cars was discussed and physically evaluated. SDM divides the sound field into spatially discrete elements of a preset analysis time window. As the native SDM assumes a wide-band source i.e. a typical full-range loudspeaker, it is normally recommended to use as short window as possible [44]. In this experiment the captured VIRs were band-limited, due to the type/size of each transducer in the cabin (Section 3.1). Hence it was possible to implement a custom-length (L) analysis window based on the properties of the transducers. For each transducer type, L was set to span three periods of the shortest wavelength in the reproduced frequency band. This allows a more accurate spatial decomposition of band-limited near-field sources - an important advantage when analyzing such complex sound fields as in car cabins.

Reproduction Protocol

SDM provides a spatial analysis and signal encoding for a given set of VIRs. The resulting data allows auralization of the sound field using a given spatial rendering scheme over loudspeakers, or Binaural Synthesis based on anechoic HRTFs. In this paper, the synthesis of the SDM-encoded spatial IRs was implemented using the *Nearest Neighbor* (NN) loudspeaker approach, similar to [57].

The performance of the system when NN is employed is highly benefited by a physical arrangement where the placement of the loudspeakers is based on the spatial analysis of the sound fields under investigation; automotive audio in this case.

Reproduction Setup

For new types of synthesized acoustic environments, designing the reproduction loudspeaker array can benefit from the spatial analysis with SDM. This approach allows the experimenter to better understand the structure of the original sound field and design an optimal reproduction layout. For this study, individual analysis of the captured VIR was employed, to ensure that the direct sound as well as reflections from the cabin's surfaces are preserved in the best possible way, during the reproduction phase. This included the aforementioned VIRs in addition to a database of twenty car cabins and system types.

The analysis followed a systematic comparison of both objective and perceptual metrics. The spatiotemporal energy distribution in time intervals [46] of each measured VIR was combined with the corresponding weighted energy error

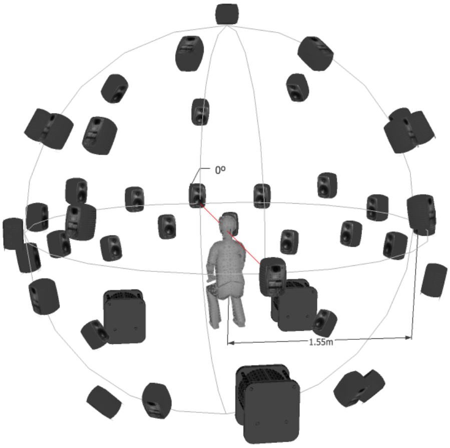


Fig. B.2: Diagram of the reproduction loudspeaker positions. The loudspeakers are placed on a spherical surface with a 1.55 m radius. The positions are left-right symmetric, and their positions (azimuth, elevation) are: 0,0; 11,-10; 22,0; 32.5,-15; 45,5; 55,-10; 65,0; 75,-10; 90,0; 120,-10; 135,10; 0,30; 40,40; 90,30; 115,30; 135,30; 150,55; 0,90; 55,-40; 120,-45; 150,-35.

estimation. This error term, results from assigning instantaneous VIR pressure to the NN loudspeakers instead of their absolute position given by the SDM analysis. The perceptual assessment focused on three perceptual constructs, *spectral fidelity*, *temporal integrity* and *accuracy of spatial representation* for two auralization sets: a full car audio system, and single transducers (see [45]). Attention was also given to the electroacoustic properties of the loudspeakers and the spatial acuity of the human hearing system, ensuring a high level of detail in the frontal plane, whilst maintaining the perceptual qualities from all directions.

For the final auralizations a 40.3 loudspeaker system was specified (see Fig. B.2). The setup comprises of 40 full-range (Genelec 8020C) and 3 Subwoofers (Genelec 7050B). One could note a second layer of loudspeakers at lower elevation (-10°) in the frontal plane (-70° to +70°). Naturally, the direct sound paths in the car cabin originate from lower elevations, as typically no transducers are placed at the ear level, and this was also evident during the spatial analysis of the VIR.

The magnitude responses of the loudspeakers in the reproduction array were confirmed in-situ, to lie primarily within ±1.5dB; including a low-frequency compensation (<200 Hz). Each loudspeaker was level-matched at the listening position, within ±0.5 dB (C-weighted RMS), using 5 s pink-noise. It is noted that the inherent inconsistencies in the physical placement of the loudspeakers introduce different times of arrival at the listening position. It was therefore required to ensure that the acoustic delay between any loudspeaker and a microphone at the listening position was temporally matched.

Reproduction Environment

When assessing spatial audio over loudspeakers, it is necessary to limit the acoustic influence of the reproduction room, on the reproduced sound field that is intended to be evaluated [37]; as it is known to be perceptible by listeners [58, 59]. This is normally achieved by ensuring that the reproduction room is characterized by a lower reverberation time compared to the room that is being reproduced. Due to the nature of the sound field in a car cabin and the very short reverberation time [29], the experimental setup used in the implementation of the current method was installed in an anechoic chamber (B5-104) located at Aalborg University. The chamber is designed and constructed to host simulation setups with human occupancy, and it is treated with absorption wedges that are 0.4 m long. Its free inner dimensions are 5.0 × 4.5 × 4.0 m. The chamber meets the requirements for anechoic performance [60] down to 200 Hz. The experimental apparatus was covered with absorption material to eliminate any reflections from the structural installation.

Visual Influence

When assessing virtual acoustics, it is important to understand and address the cross-modal behavior of the human brain. The relative importance between information within each modality (i.e. vision and hearing) is known to alter our perceptual processing [61]. This has been recognized in room acoustics research [62] in both *in-situ* and virtual acoustics assessment [63]. In a mismatch between visual and auditory cues, it has been argued that visual information may dominate the auditory sense [64]. In virtual spaces this was found to be a crucial component of user acceptance [65].

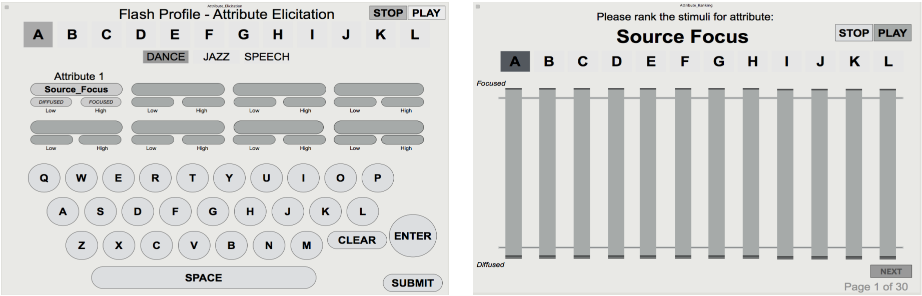
In this study we aim to reproduce a highly complex sound field which is naturally unique to car cabins, under laboratory settings. To limit possible cross-modal biases, such as visual influences, a number of steps should be followed. These controls should aim to reduce the visual influence of the experimental apparatus, to the reproduced sound field.

3.3 Evaluation - Flash Profile

The experimental design follows the principles of *Flash Profile* (FP). FP comprises of two parts: 1) the *elicitation* of attributes and 2) the *ranking* of the elicited attributes.

Flash Profile - Elicitation

During elicitation the assessor is asked to provide a non-limited number of verbal descriptors that capture the perceptual characteristics of the stimuli set under investigation. It is vital that the assessor focuses on this procedure, and define verbal descriptors that are singular, non-hedonic, and scalable, whilst they do not exhibit redundancy [37, 41].



(a) Interface during Elicitation (b) Interface during Ranking

Fig. B.3: User interface during the two phases of the experiment. Buttons labeled with letters A-L provide switching between signals. The whole experimental procedure is self-paced and self-controlled.

FP requires the whole stimuli set to be available to the assessor at all times. However, when assessing audio material the evaluation of acoustic conditions requires an excitation signal (program), i.e. music, speech, or noise [37], as each type of signal excites the conditions differently. As a consequence the stimulus is a product of a Program type (i.e. music) and a Condition (i.e. timbral alteration). To address this issue, it is recommended that during elicitation, the assessor has the option to change the program material whilst the order of the presented stimuli is maintained. This would allow assessors to explore subtle perceptual differences between specific conditions over a variety of programs, in the minimum time possible. Thus, the interface (Fig.B.3a) includes anonymously labeled stimuli (e.g. A-L) and the available program material, which assessors could select at any point during this phase. The order of presentation should be randomized between assessors.

Flash Profile - Ranking

During the ranking phase the assessor is asked to order the stimuli presented based on the perceived intensity differences of the given attribute. Each given perceptual attribute forms a block of n trials (one trial per program material). At this phase, the presentation order of the stimuli, as well as the program material, should be randomized on each trial, following typical audio evaluation conventions [37]. The graphical interface is shown in Fig. B.3b.

4 Pilot Experiment

The section above described a method for perceptual assessment of automotive audio in terms of acquiring, analyzing, reproducing, and evaluating the sound field. For reasons of completion, a brief description of a pilot study using the proposed method is included. The experiment investigated the ability of the method to assess automotive audio over basic alterations, as well as the influence of the assessor's expertise on the experimental procedure.

The example below demonstrates the perceptual effects of modifying the acoustical field in a car cabin, when altering the DSP settings as well as the physical properties of the cabin. Three system configurations were presented to the assessors: audio system without DSP processing (*No Equalization*), audio system with DSP processing (*Reference*), and audio system with DSP processing while the front side windows were open (*Front Windows Open*). The DSP processing included equalization of the transducers, delays, and individual tuning of the magnitude responses, set by an automotive sound engineer (tonmeister). The condition which included DSP processing and no physical modification of the cabin, labeled as reference, serves as the baseline and represents a premium production car audio system.

4.1 Materials & Apparatus

For this pilot experiment the car measurements were processed with the method described in Section 3.1.

Music material was then convolved with the corresponding 40.3-channel SDM responses. Here, only one program is included (*Armin van Buuren feat. Ana Criado - I'll Listen*), for simplicity. In a complete study, multiple program types should be included. The playback was based on multichannel 24-bit PCM reproduction, sampled at 48 kHz.

The assessor was given a tablet (iPad) controlling MAX 7 via MIRA remote interface, shown in Fig. B.3. The reproduction room, setup, and calibration measurements were identical to the aforementioned settings in Sec. 3.

As mentioned in Section 3.2, when reproducing sound fields in the laboratory, the visual influences of the experimental apparatus should be addressed. Here, the experiment was conducted in dark conditions and any acoustic feedback of the space was controlled. In addition, no information was given to assessors about the experimental room, the loudspeaker setup, and the content of the stimuli as recommended [26].

Assessors

Six assessors participated in the pilot experiment: two expert assessors, two experienced assessors, and two naive assessors [66]. Experts assessors ($\mathrm{As}_{1,2}$)

had more than 15 years of experience (*Mean* = 17.5) in acoustical development and critical listening; with the last 10 years focusing on automotive audio systems. Experienced assessors ($\mathrm{As}_{3,4}$) had between 3 to 6 years of experience (*Mean* = 4.5) in audio evaluation and acoustical research but no experience in automotive audio. Naive assessors ($\mathrm{As}_{5,6}$) reported no prior experience with audio evaluation or technical knowledge on the subject. All assessors were male, aged between 26-43 (*Mean* = 32, *s.d* = 6.2). The assessors' hearing sensitivity was confirmed with standard procedures [67] to be above 20 dBHL at 125-8000 Hz.

In the case where assessors were not familiar with verbal elicitation procedures, an introduction was given and the assessors performed verbal elicitation for a set of visual stimuli.

4.2 Results

Figure B.4 presents the perceptual responses, i.e., the sensory profiles, of all assessors for the three acoustical conditions used in the evaluation. Overall, it can be seen that all assessors identified both timbral effects and spatial differences between the presented stimuli and created similar profiles. Remarkable timbral differences can be seen between the conditions of no equalization and reference. Similarly, alterations of spatial properties were identified for the condition where the windows were open. Opening the front windows revealed less prominent timbral effects, compared to no equalization, yet, they were identified and rated accordingly by most assessors.

One should note that in FP, the elicited individual vocabulary of each assessor is used as given by each assessor, in contrast with consensus methods were a common vocabulary is defined and used by all assessors. Thus, the semantic meaning of each attribute used in FP, is unique to each assessor. It is expected that the scaling, the extreme intensities (scale anchors), and the perceptual constructs underlying the ratings may differ across individuals. That makes the direct comparison across assessors a non trivial matter, even when the same label was used [68]. To limit possible ambiguities of attributes' labels and identify relationships between descriptors given by different assessors, each assessor provided definitions and anchors for their attributes. The definitions are summarized in Appendix (Table B.2).

Here, a limited number of attributes is selected for comparison, where their definitions and anchors indicate strong relationships across assessors, and tangible relation to physical quantities was possible (i.e. low frequency alterations). This would allow a basic comparison of individually elicited attributes across multiple assessors, in order to provide holistic insights within this exploratory investigation.

As shown in Fig. B.4, no equalization condition seem to strongly affect the perceived bass content, indicating lower intensities for $\mathit{Bass\ Quantity}_{\mathrm{As1}}$,

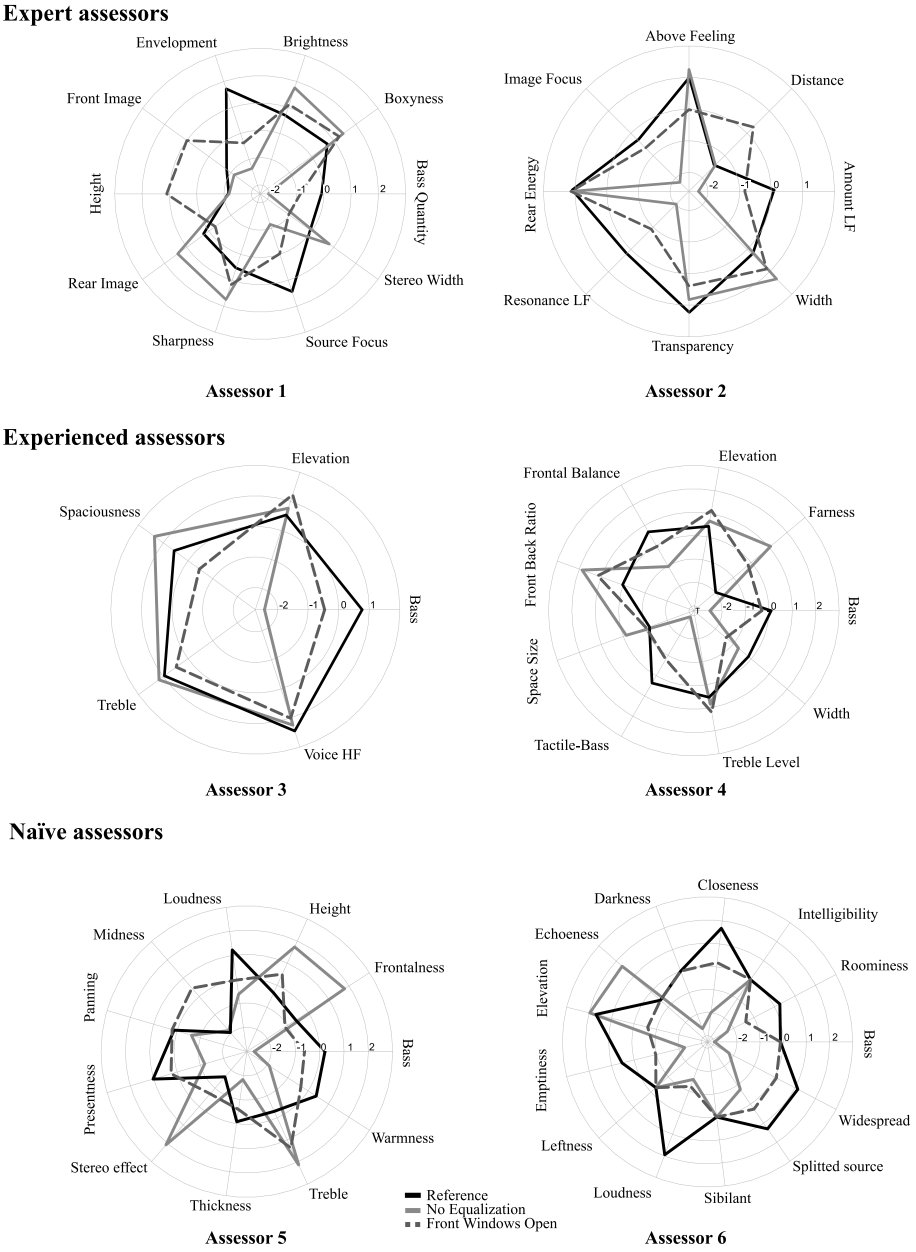
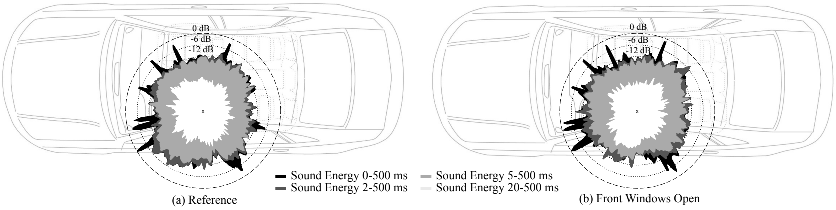


Fig. B.4: Sensory profiles depicting the assessors' responses for the three acoustical conditions. To limit scaling effects between the assessors and allow comparisons the raw data were centered and normalized (*Mean* = 0). The attributes' definitions are given in Table B.2.



(a) Reference

(b) Front Windows Open

Fig. B.5: Spatiotemporal visualization [46] based on the directional energy response of the system, for Reference (a) and Front Windows Open settings (b).

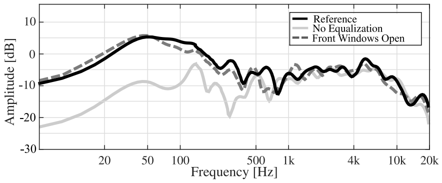


Fig. B.6: The total magnitude response of the car audio system at the listening position for (1) Reference, (2) No Equalization, and (3) Front Windows Open conditions.

Amount LF$_{\mathrm{As2}}$, *Resonance LF*$_{\mathrm{As2}}$, *Bass*$_{\mathrm{As3,4,5,6}}$, and *Tactile Bass*$_{\mathrm{As4}}$, compared to the reference. Less prominent alterations at low frequencies were identified by assessors for front windows open. All assessors, independent of training experience, have equally identified these differences at the appropriate intensity levels. The perceptual data come in agreement with the known physical changes between these two conditions, shown in Figure B.6; compared to reference, no equalization condition included differences up to 20 dB, whilst front windows open condition was limited to 3 dB in that frequency range.

Further, assessors indicated that no equalization has a negative effect on attributes related to the feeling of sound arriving from multiple directions e.g., *Envelopment*$_{\mathrm{As1}}$, *Spaciousness*$_{\mathrm{As3}}$, and the ability to locate the sound source, given by *Source Focus*$_{\mathrm{As1}}$, *Image Focus*$_{\mathrm{As2}}$,and *Frontal Balance*$_{\mathrm{As4}}$. The source's positioning and focus was also indicated as alteration of the perceived width, e.g., on *Stereo Width*$_{\mathrm{As1}}$, *Width*$_{\mathrm{As2}}$, *Stereo Effect*$_{\mathrm{As5}}$, *Splitted Source*$_{\mathrm{As6}}$ and *Widespread*$_{\mathrm{As6}}$. Moreover, assessors identified noticeable differences for attributes relating to the ratio between the energy coming from the front and back directions, such as *Rear Image*$_{\mathrm{As1}}$, *Front Back Ratio*$_{\mathrm{As4}}$, and *Frontalness*$_{\mathrm{As5}}$. The above perceived differences across the assessors indicate spectral and level imbalances in the no equalization condition, compared to that of the reference. The identified differences are amongst the common characteristics that sound engineers aim to improve using DSP processing, as in the presented reference condition. One should expect such perceptual effects to be altered in the absence of DSP processing.

Based on previous investigations (see [20]), altering the front side windows is likely to affect perceptual attributes relating to *Apparent Source Width* and *Distance*. Due to changes in the volumetric properties of the cabin [13], the low frequency behavior should also change. Figure B.5 shows the spatiotemporal analysis of reference condition and when the front windows were opened. It can be seen that the energy originating from sides (±60°) is decreased when the windows are open, for the first three time intervals. It also noted that the overall reproduction symmetry of the system would be affected based on the depicted energy distribution after the first 20 ms.

The perceived differences (Fig. B.4) seem to depict similar observations, as opening the front windows reveals high intensity on spatial attributes such as the perceived $\textit{Front Image}_{\mathrm{As6}}$, $\textit{Stereo Width}_{\mathrm{As1}}$ and $\textit{Width}_{\mathrm{As2,4}}$, compared to the reference condition. Timbral effects have been also noted by most assessors, indicating decreased bass content. Interestingly, although opening the windows shows changes on the perceived $\textit{Height}_{\mathrm{As1}}$, $\textit{Above Feeling}_{\mathrm{As2}}$, and $\textit{Distance}_{\mathrm{As2}}$, removing the equalization reveals no such differences. This follows the current understanding, as the perceived distance is known to relate to lateral reflections [20], and such differences could not be easily elicited by simple equalization settings.

Table B.1 reports the explained variances per dimension for each assessor, using *Principal Component Analysis* (PCA) [70]. The more trained assessors were able to identify the perceptual differences, and rate them with attributes that support both dimensions (Dim. 1 $\approx$ 65%, Dim. 2 $\approx$ 35%). The judgments of less trained assessors were primarily explained by the first dimension ($\approx$ 85%), and much lower second dimension ($\approx$ 15%). To further analyze this finding, the correlation of the attributes to the identified dimensions was also calculated, given as the *Correlation Ratio* (CR). This metric effectively indicates the quality of the given attributes and the ability of the assessors to quantify and differentiate the perceived sensations in the multidimensional space. The CR for the second dimension is marginally lower for naive assessors, than the experts and the experienced. That denotes the enhanced ability of the trained assessors to conduct the experiment reliably, with well supported attributes, as noted before [54].

In terms of the assessors efficiency, the current observations suggest that the more trained assessors were able to perform the experiment in less time, as shown in Fig. B.7. Experts spent less time during the elicitation procedure compared to the less trained assessors. This confirms the advantage of using trained assessors in such protocols, due to their enhanced ability to efficiently identify and verbalize perceived differences, in the auditory domain.

Overall, the current results indicate that the assessors perceived the physical changes in an expected way and they are in close agreement with previous elicitation studies in automotive audio [12] and spatial audio reproduction [20]. It should be noted that the data presented here is an illustrative set to initially assess the method proposed, and should not be used to conclude findings about

As	Dimension 1	Dimension 2	CR $_{\text{Dim.1}}$	CR $_{\text{Dim.2}}$
1	67.03	32.97	80	50
2	61.88	38.12	75	50
3	70.74	29.26	80	40
4	68.83	31.17	89	44
5	84.60	15.40	91	9
6	81.90	18.10	69	15

Table B.1: Results of Principal Component Analysis for each assessor. Correlation Ratio (CR) refers to the percentage of the well correlated attributes (R > |0.5|) to a dimension, noted as (C), divided by the total number of given attributes of that assessor (T) i.e. $\frac{C}{T}$ ×100.

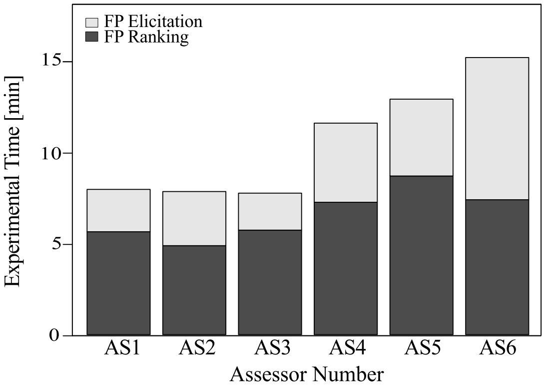


Fig. B.7: Time required for each assessor to perform the elicitation and ranking phases of Flash Profile.

car acoustics, due to the limited contextual factors. An in-depth analysis of car acoustics and the experimental setup should follow this preliminary investigation, to further validate and establish the experimental protocol applied.

5 Discussion

Although experiencing automotive sound in-situ is unequivocally a true reference to reality, in practice it could be employed in a limited number of scientific investigations, e.g., comparing different equalization and signal processing algorithms within a cabin. Expanding in-situ protocols to investigations of physical parameters (i.e. loudspeaker placement, cabin acoustics) is impractical and highly laborious. Even if made possible, it would inherently include long test-to-test periods, contrasting the basic requirements of perceptual evaluation of audio material.

Here, a laboratory-based protocol was proposed. SDM was employed for the acquisition and presentation of automotive audio. This approach maintains the benefits of perceptual assessment in the laboratory similar to the aforementioned methods (Sec. 2.2). It therefore allows instant, double-blind, and comparative assessment of different sound fields in the laboratory. Yet, it overcomes issues related to non-auditory feedback (i.e. brand) during *in-situ* evaluation and the limitations imposed from headphone-based playback of BCS.

Dealing with perception requires to employ a scientific evaluation method where the perceptual properties are quantified, whilst it avoids uncontrolled biases on human's sentiments and judgments, such as individual's taste, beliefs, experience, and needs [37, 69]. Descriptive Sensory Analysis (DA) methodologies, found in food and wine industry [41], have been instrumental in decoding such complex perceptual constructs. Still, they are laborious and require long experimental procedures. Here, we propose the use of FP, a rapid sensory analysis technique, developed to provide a sensory profiling similar to common DA, still at the least time possible (e.g. typically within a 1-3h session).

Multiple limitations exist within automotive audio evaluation which FP seem to overcome. First, the time restrictions imposed will be well addressed when using FP, yet the benefits of conducting DA are preserved. Moreover its need of only 4-5 (expert) assessors with general sensory expertise, instead of higher numbers of product-trained assessors [26], its comparative ranking nature that does not require audio reference, as well as the flexibility of the method may well suit the processes within automotive audio.

In fact, the comparative nature of FP combined with direct access to the entire stimuli set simultaneously, allows the assessors to adapt their cognitive strategies during the evaluation. When performing FP, the assessors may employ both comparative and short memory-based judgments. That results to a higher number of direct comparisons than traditional forced-choice comparison methods [26, 53]. It also permits the evaluation of very similar, or highly dissimilar stimuli within the same procedure, with no statistical disadvantage, as common repeated measures protocols [37].

In the pilot study presented, the performance of experts, experienced, and naive assessors was investigated. The results indicate that experts completed the experimental procedure faster (Fig. B.7), providing precise, and technically familiar attributes which correspond well to the subsequent ratings of the stimuli (Table B.1), as well as to their definitions (Table B.2).

The quality and number of non-correlated attributes is vital requirement for a fast and efficient profiling during FP [26]. Based on the current findings, one could argue that naive listeners failed to meet this requirement, as they provided a high number of inter-correlated attributes, some of which did not contribute to the explained variances, as no differences were identified. The two experts, provided clear results, even for physically small differences, e.g., the low frequency alteration between reference and front windows open. These alterations were identified also by experienced and naive listeners, but not in a consistent manner, as noted in previous investigations of FP in food products [54].

Moreover, the nature of FP highly benefits from a common background of auditory sensitivity across assessors, so that the perceptual constructs driving the ratings could be used in a common factorial space [26]. This is an important element when performing multitable quantitative analysis, as the typical FP statistical analysis processes. Here, the results illustrate similarities between assessors of similar experience level, but not as well across them. The PCA analysis (Table B.1) indicates that although all assessors performed adequately the task, a strong common statistical relationship across all assessors, might be difficult to achieve. It is therefore recommended that further investigations should consider assessors' background when establishing an evaluation panel for FP.

Finally FP is based on quantitative description. Here, a simple data analysis was presented for completeness. Although such an analysis is informative, the real benefit of FP is realized when it is combined with multivariate statistics, such as *Multiple Factor Analysis* [71]. This type of analysis allows the investigation of both the individually elicited attributes as well as the given ratings in a common factorial space for all assessors. Effectively, FP merges quantitative and qualitative data using a mathematical approach rather than subjective analysis of the experimenter [70].

5.1 Limitations & Future work

One should note that the SDM provides a faithful and plausible acoustical representation, however, as any spatial reproduction method to date, it has certain limitations. It was shown recently [45] that a post-equalization of the analyzed response is needed when SDM is applied to cars due to high echo density. There it was also shown that the complex geometry of a car cabin and the extreme acoustical conditions may violate the basic assumptions of SDM of plane waves. As an *Impulse Response* (IR) -based method, several aspects of the field are not captured during acquisition, i.e. the non-linearities or structure-bone vibrations, as well as the acoustical effects of the human body. Thus, one should not expect that the reproduced sound field is an exact physical replica of the real field.

Nevertheless, both objective and perceptual results suggest that SDM preserves the basic perceptual differences between the stimuli set in the pilot experiment. The results of the pilot study support the capacity of the method to address both timbral and spatial properties of the sound fields in cars. Further advantages of the method include the flexible reproduction scheme, the fast acquisition of VIR compared to BCS, as well as the analysis and novel visualization capabilities of the spatial properties of the sound field.

Rapid Sensory Analysis methods aim to remove tedious processes within

DA, such as building a consensus vocabulary of well-defined descriptors across assessors, as well as several training and evaluation sessions. As a direct consequence, when employing FP the experimenter cannot argue easily on the semantic meaning of the descriptors. Moreover, FP gains experimental time by not including repeats or hidden anchors. For audio material, this is a serious limitation as it is well known that the context and properties of the program material influence the results [37]. In fact critical listeners often require specific program material for assessing certain attributes (e.g. speech content for *intelligibility* ratings). Therefore, when conducting FP of audio it is strongly recommended to employ several program materials.

Further work on this topic includes the perceptual assessment of automotive audio systems and car cabins in detail, aiming to identify the perceptual constructs originating from acoustical alterations in the sound field in question. Additional validation studies of the proposed method should be conducted, for example by comparing the results to in-situ evaluation when possible, as well as contrasting FP to common DA procedures with identical contextual factors. Such approaches would improve the understanding of the method and further validate its applicability in audio evaluation.

Here, the reproduction of the auralized stimuli is conducted over loudspeakers in an anechoic chamber. It is noted that the presentation and evaluation method could be altered to meet further research objectives and the related practical implications. For example, the presentation on headphones is still possible, as well as following standardized audio evaluation methods. Yet, the spatiotemporal analysis of the measured field will still be available. Moreover, the experimental methodology proposed here could be applied in several domains, for example in assessing room acoustics. Future work could perceptually assess a variety of acoustical settings in standard everyday rooms, to better understand the influence of the acoustical properties of a reproduction room that are inherently imposed on the reproduced sound field.

6 Concluding Remarks

This paper reviewed the past and current practices for perceptual assessment of automotive audio and stipulated new approaches and research paths to address the industry's future demands. A novel method was described in terms of acquiring, presenting, and evaluating the reproduced sound of automotive systems, targeting a general framework for perceptual studies in automotive audio. Finally, a pilot experiment was presented and the preliminary results were discussed.

The method applied SDM for capturing, analyzing and presenting the sound field to human assessors. The current results, indicate that this approach yields faithful representation of timbral and spatial properties of automotive sound, whilst providing novel analysis tools. Employing this protocol may allow several properties of automotive audio to be assessed in controlled and instantaneous comparative protocols, including both acoustical, as well as electrical changes over the same experimental paradigm.

Conducting the evaluation using rapid sensory analysis such as Flash Profile, allows the assessors to use their own vocabulary to describe and quantify the auditory sensations within a single experimental method. Thus, such approach may permit a more detailed assessment of unique and novel experiences in cars, e.g. 3D sound reproduction and upmixing, noise masking schemes, and individual sound zones within cars.

Acknowledgments

The authors would like to thank Morten Lydolf, Martin Møller, Martin Olsen, Claus Vestergaard Skipper, and their colleagues at Aalborg University for their support and helpful input. The research leading to these results has received funding from the European Union's Seventh Framework Program under grant agreement no. ITN-GA-2012-316969.

7 Appendix

Attribute	Low anchor	High anchor	Definition
Bass Quantity$_{\mathrm{As1}}$	too little	too much	level of bass content
Boxyness$_{\mathrm{As1}}$	boxy	open	some sound boxy, in a very small space
Brightness$_{\mathrm{As1}}$	dull	bright	how bright is the sound (high frequencies)
Envelopment$_{\mathrm{As1}}$	low	high	how enveloping is the sound field, sound from around
Front Image$_{\mathrm{As1}}$	side	front	if the frontal image is exactly in front of me
Height$_{\mathrm{As1}}$	absent	present	sound appears higher points
Rear Image$_{\mathrm{As1}}$	absent	present	sometimes sound appears from the back
Sharpness$_{\mathrm{As1}}$	blunt	sharp	sharp/transients at mid-high freq.
Source Focus$_{\mathrm{As1}}$	diffuse	focused	sourced located at a point?
Stereo Width$_{\mathrm{As1}}$	narrow	wide	how wide the sound image is
Above Feeling$_{\mathrm{As2}}$	low	high	feeling that sound comes from above me
Amount LF$_{\mathrm{As2}}$	none	a lot	the low frequency energy levels
Distance$_{\mathrm{As2}}$	near	far	how distant the image feels
Image Focus$_{\mathrm{As2}}$	muffled	focused	image clarity / clearly at a location
Rear Energy$_{\mathrm{As2}}$	low	high	the amount of sound energy coming from the back
Resonant LF$_{\mathrm{As2}}$	not resonant	resonant	specific low frequency resonance
Transparency$_{\mathrm{As2}}$	not transparent	transparent	how transparent the sound is
Width$_{\mathrm{As2}}$	narrow	wide	the width of the sound image
Bass $_{\mathrm{As3}}$	low	high	how much bass there is (spectrum)
Elevation$_{\mathrm{As3}}$	low	high	if the sound appears higher than 0°
Spaciousness$_{\mathrm{As3}}$	frontal	lateral	if the sound comes from front or multiple lateral directions
Treble$_{\mathrm{As3}}$	low	high	how much treble there is (spectrum)
Voice HF$_{\mathrm{As3}}$	muffled	bright	if the vocal is bright, as present
Bass$_{\mathrm{As4}}$	low	high	bass levels
Farness$_{\mathrm{As4}}$	near	far	Distance - the source is not as present
Elevation$_{\mathrm{As4}}$	lower	higher	sound perceived as higher (location) as others
Front Back Ratio$_{\mathrm{As4}}$	too back	too front	the level (balance) of the sound coming from back and front
Frontal Balance$_{\mathrm{As4}}$	too left	too right	the level (balance) of the stereo image
Space Size$_{\mathrm{As4}}$	small	big	how big does the room that I am in, feels
Tactile Bass$_{\mathrm{As4}}$	not there	there	a feeling of low frequency pressure or body vibration
Treble Level$_{\mathrm{As4}}$	little	too much	the high frequency content level
Width$_{\mathrm{As4}}$	toowide	focused	the horizontal size of the ensemble
Bass$_{\mathrm{As5}}$	a little	a lot	how much bass I hear
Frontalness$_{\mathrm{As5}}$	front	back	balance of front-back sound
Height$_{\mathrm{As5}}$	little	a lot	sound from above, higher than others
Loudness $_{\mathrm{As5}}$	low	high	level differences
Midness$_{\mathrm{As5}}$	side	mid	sound from centre (front) or sides
Panning $_{\mathrm{As5}}$	left	right	imbalance of left-right sound
Presentness$_{\mathrm{As5}}$	not present	present	presence, prominent sound
Stereo Effect$_{\mathrm{As5}}$	a little	a lot	sound from sides, like wide stereo
Thickness$_{\mathrm{As5}}$	thin	thick	feeling of thick, full sound
Treble$_{\mathrm{As5}}$	low	high	higher frequencies level
Warmness $_{\mathrm{As5}}$	cold	warm	feeling for sound being warm
Bass$_{\mathrm{As6}}$	low	high	bass level
Closeness$_{\mathrm{As6}}$	not close	close	does the sound feel close to me?
Darkness$_{\mathrm{As6}}$	bright	dark	dark (frequencies) sounds or not
Echoeness$_{\mathrm{As6}}$	not much	too much	if I can hear echo
Elevation$_{\mathrm{As6}}$	ok	high	playback from higher positions
Emptiness$_{\mathrm{As6}}$	empty	full	the room sounds empty
Intelligibility$_{\mathrm{As6}}$	not	ok	quality of voice to understand, like bad codec
Leftness$_{\mathrm{As6}}$	too left	center	more sound from left
Loudness$_{\mathrm{As6}}$	lower	higher	the level of the music
Room Print$_{\mathrm{As6}}$	there	not there	how loud is the room in sound
Sibilant$_{\mathrm{As6}}$	ok	too much	‘s’ or ‘z’ sound in speech
Splitted source$_{\mathrm{As6}}$	split	one point	sound from many directions or not
Widespread$_{\mathrm{As6}}$	around	front	sound spread around

Table B.2: Individually elicited attributes, including the provided anchors and definitions. The attributes shown were elicited using a larger stimuli set.

References

[1] F. Rumsey, "Automotive Audio," *J. Audio Eng. Soc.*, vol. 60, no. 12, pp. 1070–1074 (2012).

[2] IEC, "60268-13 (BS 6840-13) Sound system equipment - Part 13: Listening tests on loudspeakers," International Electrotechnical Commission, Tech. Rep. (1998).

[3] ITU-R Rec., "BS 1534-1: Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)," International Telcommunication Union Recommendations, Geneva, Switzerland, Tech. Rep. (2001).

[4] M. Strauss, J. Nowak, and D. de Vries, "Approach to sound field analysis and simulation inside a car cabin," in *Proc. 36th Int. Conf. Audio Eng. Soc. Automot. Audio*, Dearborn, Michigan, 2009.

[5] M. Kleiner and C. Lindgren, "Objective Characterization Automotive of Audio Sound Spaces Fields in," in *Proc. 15th Int. Conf. Audio Eng. Soc.*, 1998.

[6] E. Granier, M. Kleiner, B. Dalenbäck, and P. Svensson, "Experimental Auralization of Car Audio Installations," *J. Audio Eng. Soc.*, vol. 44, no. 10, pp. 835–849 (1996).

[7] ISO, "3382-1: Acoustics Measurement of Room Acoustic Parameters Part 1: Performance Spaces," International Standards Organisation, Geneva, Switzerland, Tech. Rep. (2009).

[8] C. Choi, L.-h. Kim, S. Doo, Y. Oh, and K.-m. Sung, "Assessment of Sound Field in a CAR," in *113rd Conv. Audio Eng. Soc.*, Los Angeles, USA, 2002, convention paper #5701.

[9] L. Klapproth and B.-w. Gmbh, "Acoustic Characteristics of the Vehicle Environment," in *77th Conv. Audio Eng. Soc.*, Hamburg, Germany, 1985, convention paper #2185.

[10] G. Martin and S. Bech, "Attribute Identification and Quantification in Automotive Audio - Part 1: Introduction to the Descriptive Analysis Technique," in *118th Conv. Audio Eng. Soc.*, Barcelona, Spain, 2005, convention paper #6360.

[11] S. Bech, M.-A. Gulbol, G. Martin, J. Ghani, and W. Ellermeier, "A listening test system for automotive audio – Part 2- Initial verification," in *118th Conv. Audio Eng. Soc.*, Barcelona, Spain, 2005, convention paper #6359.

[12] P. Hegarty, S. Choisel, and S. Bech, “A Listening Test System for Automotive Audio - Part 3: Comparison of Attribute Ratings Made in A Vehicle with those made using an Auralization System,” in *123rd Conv. Audio Eng. Soc.*, New York, USA, 2007, convention paper #7224.

[13] F. Christensen, G. Martin, M. Lydolf, M. Pauli, S. Woo-Keun, and P. Benjamin, “A Listening Test System for Automotive Audio - Part 1 : System Description,” in *118th Conv. Audio Eng. Soc.*, Barcelona, Spain, 2005, convention paper #6358.

[14] R. Shively, “Subjective Evaluation of Reproduced Sound in Automotive Spaces,” in *Proc. 15th Int. Conf. Audio Eng. Soc.*, Copenhagen, Denmark, 1998.

[15] N. Otto, “Listening Test Methods for Automotive Sound Quality,” in *103rd Conv. Audio Eng. Soc.*, New York, USA, 1997, convention paper #4586.

[16] A. Azzali, A. Farina, G. Rovai, G. Boreanaz, and G. Irato, “Construction of a Car Stereo Audio Quality,” in *117th Conv. Audio Eng. Soc.*, San Fransisco, USA, 2004, convention paper #6306.

[17] E. Ugolotti, G. Gobbi, and A. Farina, “IPA - A Subjective Assessment Method of Sound Quality of Car Sound Systems,” in *110th Conv. Audio Eng. Soc.*, Amsterdam, The Netherlands, 2001,convention paper #5281.

[18] S. E. Olive, “A New Reference Listening Room For Consumer, Professional and Automotive Audio Research,” in *126th Conv. Audio Eng. Soc.*, Munich, Germany, 2009, convention paper #7677.

[19] W. N. House, “Aspects of the Vehicle Listening Environment,” in *87th Conv. Audio Eng. Soc.*, New York, USA, 1989, convention paper #2873.

[20] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, “Perception of Reverberation in Small Rooms : A Literature Study,” in *Proc. 55th Int. Conf. Audio Eng. Soc. Spat. Audio*, Helsinki, Finland, 2014.

[21] S. E. Olive and T. Welti, “Validation of a Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems,” in *Proc. 36th Int. Conf. Audio Eng. Soc. Automot. Audio*, 2009, pp. 1–7.

[22] F. E. Toole and S. E. Olive, “Hearing is believing vs. believing is hearing: blind vs. sighted listening tests, and other interesting things,” *Proc. 97th Conv. Audio Eng. Soc.*, vol. 3894 (1994).

[23] N. House and R. Shively, “The Placebo Method: A comparison of the In-Situ Subjective Evaluation Methods For Vehicles,” in *108th Conv. Audio Eng. Soc.*, Paris, France, 2000, convention paper #5136.

[24] S. Cecchi, A. Primavera, F. Piazza, F. Bettarelli, E. Ciavattini, R. Toppi, J. G. F. Coutinho, W. Luk, C. Pilato, F. Ferrandi, V. M. Sima, and K. Bertels, "The hArtes CarLab: A new approach to advanced algorithms development for automotive audio," *J. Audio Eng. Soc.*, vol. 59, no. 11, pp. 858–869 (2011).

[25] M. Sams, R. Hari, J. Rif, and J. Knuutila, "The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence," *J. Cogn. Neurosci.*, vol. 5, no. 3, pp. 363–368 (1993).

[26] J. Delarue, B. J. Lawlor, and M. Rogeaux, *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research.* Elsevier, 2014.

[27] A. J. Svobodnik and G. Ruiz, "Virtual Development of Mercedes Premium Audio Systems," in *Proc. 36th Int. Conf. Audio Eng. Soc. Automot. Audio*, 2009, pp. 1–6.

[28] N. Harris, "A comparison of modelling techniques for small acoustic spaces such as car cabins," in *122nd Conv. Audio Eng. Soc.*, Vienna, Austria, 2007, convention paper #7146.

[29] E. Granier, M. Kleiner, B. Dalenbäck, and P. Svensson, "Experimental Auralization of Car Audio Installations," *J. Audio Eng. Soc.*, vol. 44, no. 10, pp. 835–849 (1996).

[30] S. Choisel, P. Hegarty, F. Christensen, B. Pedersen, and W. Ellermeier, "A listening test system for Automotive Audio - Part 4: Comparison of attribute ratings made by expert and non-expert listeners," in *123rd Conv. Audio Eng. Soc.*, New York, USA, 2007, convention paper #7225.

[31] P. Mackensen, U. Felderhof, G. Theile, U. Horbach, and R. Pellegrini, "Binaural room scanning - A new tool for acoustic and psychoacoustic research," *J. Acoust. Soc. Am.*, vol. 105, no. 2, p. 1343 (1999). doi: 10.1121/1.426373

[32] A. Roginska, G. H. Wakefield, and T. Santoro, "User Selected HRTFs: Reduced Complexity and Improved Perception," in *Proc. Undersea Hum. Syst. Integr. Symp.*, Providence, IL, 2010.

[33] J. Blauert, *The Technology of Binaural Listening*, J. Blauert, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[34] W. Rudmose, "The case of the missing 6 dB," *J. Acoust. Soc. Am.*, vol. 71, no. 3, p. 650 (1982). doi: 10.1121/1.387540

[35] W. L. Martens, H. Sakanashi, and W. Woszczyk, “Whole-body vibration associated with low-frequency audio reproduction influences preferred equalization,” in *Proc. 36th Int. Conf. Audio Eng. Soc. Automot. Audio*, Dearborn, USA, 2009.

[36] S. Germain, S. E. Olive, and T. Welti, “The Effect of Whole-body Vibrations on Preferred Bass Equalization of Automotive Audio Systems,” in *127th Conv. Audio Eng. Soc.*, New York, USA, 2009, convention paper #7956.

[37] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, method and application.* Wiley, 2007.

[38] P. Dennis, “In-Vehicle Audio System Sound Quality Preference Study,” in *Proc. 139th Conv. Audio Eng. Soc.*, New York, 2015, pp. 1–12.

[39] A. Farina and E. Ugolotti, “Subjective Comparison of Different Car Audio Systems by the Auralization Technique,” in *Proc. 103rd Conv. Audio Eng. Soc.*, New York, USA, 1997.

[40] D. Clark, “Listening Technology for Autmotive Sound Systems,” in *114th Conv. Audio Eng. Soc.*, Amsterdam, The Netherlands, 2003, convention paper #5723.

[41] H. T. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices.* New York: Springer, 1999.

[42] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, “Elicitation of attributes for the evaluation of audio-on-audio interference,” *J. Acoust. Soc. Am.*, vol. 136, no. 5, pp. 2630–2641 (2014). doi: 10.1121/1.4898053

[43] A. Melka and P. Pelant, “Psychoacoustic methods applied to SKODA passenger car development programmes: intentions and first experiences,” *Acta Acust. united with Acust.*, vol. 85, pp. 657–660 (1999).

[44] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, “Spatial decomposition method for room impulse responses,” *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 17–28 (2013).

[45] S. Tervo, J. Pätynen, N. Kaplanis, M. Lydolf, S. Bech, and T. Lokki, “Spatial Analysis and Synthesis of Car Audio System and Car-Cabin Acoustics with a Compact Microphone Array,” *J. Audio Eng. Soc*, vol. 63, no. 11, pp. 914–925 (2015).

[46] J. Pätynen, S. Tervo, and T. Lokki, “Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses.” *J. Acoust. Soc. Am.*, vol. 133, no. 2, pp. 842–57 (2013). doi: 10.1121/1.4770260

[47] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, “Disentangling preference ratings of concert hall acoustics using subjective sensory profiles,” *J. Acoust. Soc. Am.*, vol. 132, no. 5, pp. 3148–3161 (2012).

[48] A. Kuusinen, J. Pätynen, S. Tervo, and T. Lokki, “Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls.” *J. Acoust. Soc. Am.*, vol. 135, no. 1, pp. 239–250 (2014). doi: 10.1121/1.4836335

[49] J. Berg and F. Rumsey, “Spatial attribute identification and scaling by repertory grid technique and other methods,” in *Proc. 16th Int. Conf. Audio Eng. Soc. Spat. Audio*, Rovaniemi, Finland, 1999.

[50] G. Lorho, “Perceived Quality Evaluation - An application to sound reproduction over headphones,” Ph.D. dissertation, Aalto University (2010).

[51] C. Simonsen and S. Legarth, “A procedure for sound quality evaluation of hearing aids,” *Hear. Rev.*, vol. 17, no. 13, pp. 32–37 (2010).

[52] N. Zacharov, J. Ramsgaard, G. Le Ray, and C. V. Jorgensen, “The multidimensional characterization of active noise cancelation headphone perception,” in *Second Int. Work. Qual. Multimed. Exp.*, 2010, pp. 130–135.

[53] V. Dairou and J.-M. Sieffermann, “A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile,” *J. Food Sci.*, vol. 67, no. 2, pp. 826–834 (2002). doi: 10.1111/j.1365-2621.2002.tb10685.x

[54] J. Delarue and J.-M. Sieffermann, “Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products,” *ood Qual. Prefer.*, vol. 15, no. 2, pp. 383–292 (2004).

[55] S. Parkin, G. M. Mackay, and A. Cooper, “How Drivers Sit in Cars,” *Accid. Anal. Prev.*, vol. 27, no. 6, pp. 777–783 (1995).

[56] A. Farina, “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique,” in *108th Conv. Audio Eng. Soc.* Paris, France: Audio Engineering Society, feb 2000, convention paper #5093.

[57] S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki, “Preferences of critical listening environments among sound engineers,” *J. Audio Eng. Soc.*, vol. 62, no. 5, pp. 300–314 (2014).

[58] F. Rumsey, “Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm,” *J. Audio Eng. Soc.*, pp. 651–666 (2002).

[59] F. E. Toole, *Sound Reproduction: Loudspeakers and Rooms.* Boston: Focal Press, 2008.

[60] ISO, "3745 - Determination of sound power levels and sound energy levels of noise sources using sound pressure. Precision methods for anechoic rooms and hemi-anechoic rooms," International Organization for Standardization, Tech. Rep. (2003).

[61] B. E. Stein and M. A. Meredith, *The Merging of the Senses.* Boston: MIT Press, 1993.

[62] A. Defays, S. Safin, A. Billon, C. Decaestecker, N. Warzée, P. Leclercq, and A.-S. Nyssen, "Bimodal Interaction: The Role of Visual Information in Performing Acoustic Assessment in Architecture," *Ergon. Open J.*, vol. 7, pp. 13–20 (2014).

[63] C. Nathanail, C. Lavandier, J.-D. Polack, and O. Warusfel, "Influence of sensory interactions between vision and audition on the perceptual characterization of room acoustics," *J. Acoust. Soc. Am.*, vol. 105, no. 2, p. 1046 (1999). doi: 10.1121/1.424982

[64] P. Larsson and D. Västfjäll, "When what you hear is what you see: Presence and auditory-visual integration in virtual environments," in *Proc. 10th Annu. Int. Work. Presence*, Barcelona, Spain, 2007, pp. 11–18.

[65] P. Svensson, "On Reverberation Enhancement in Auditoria," Ph.D. dissertation, Chalmers University of Technology (1994).

[66] ITU-R Rec., "BS 2300-0: Methods for Assessor Screening," International Telecomm Union, Geneva, Switzerland, Tech. Rep. (2014).

[67] ISO, "BS 8253-1: Audiometric test methods Part 1: Pure-tone air and bone conduction," International Organization for Standardization (2010).

[68] J. Pagès, *Multiple Factor Analysis by Example Using R.* CRC Press, 2014, vol. 20.

[69] V. Hulusic, C. Harvey, K. Debattista, N. Tsingos, S. Walker, D. Howard, and A. Chalmers, "Acoustic Rendering and Auditory-Visual Cross-Modal Perception and Interaction," *Comput. Graph. Forum*, vol. 31, no. 1, pp. 102–131 (2012). doi: 10.1111/j.1467-8659.2011.02086.x

[70] S. Lê and T. Worch, *Analyzing Sensory Data with R.* Boca Raton, FL: CRC Press, 2015.

[71] H. Abdi, L. J. Williams, and D. Valentin, "Multiple factor analysis: principal component analysis for multitable and multiblock data sets," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 5, no. 2, pp. 149–179 (2013). doi: 10.1002/wics.1246

Paper C

Perceptual Aspects of Reproduced Sound in Car Cabin Acoustics

Neofytos Kaplanis, Søren Bech, Sakari Tervo, Jukka Pätynen,
Tapio Lokki, Toon van Waterschoot, and Søren Holdt Jensen.

The paper has been accepted for publication in the
Journal of Acoustical Society of America accepted, 2017.

The layout has been revised.

Abstract

An experiment was conducted to determine the perceptual effects of car cabin acoustics on the reproduced sound field. In-car measurements were conducted whilst the cabin's interior was physically modified. The captured sound fields were recreated in the laboratory using a 3D loudspeaker array. A panel of expert assessors followed a rapid sensory analysis protocol, the Flash Profile, to perceptually characterize and evaluate twelve acoustical conditions of the car cabin, using individually elicited attributes. A multivariate analysis revealed the panel's consensus and the identified perceptual constructs. Six perceptual constructs characterize the differences between the acoustical conditions of the cabin, related to Bass, Ambience, Transparency, Width & Envelopment, Brightness, and Image focus. The current results indicate the importance of several acoustical properties of a car's interior on the perceived sound qualities. Moreover, they signify the capacity of the applied methodology in assessing spectral and spatial properties of automotive environments in laboratory settings, using a time-efficient and flexible protocol.

1 Introduction

Automotive environments are steadily becoming popular listening spaces. Aiming towards a high quality reproduction, in–car audio systems have reformed from an adequate monophonic reproduction at first, to today's multichannel loudspeaker systems capable of delivering some of the most advanced audio schemes available [1]. In acoustical terms, automotive audio systems exhibit unique and domain-specific challenges that increase the complexity and the development requirements [2, 3]. The unconventional and adverse acoustical properties of the car cabins [3, 4] are unequivocally the dominant challenges when developing such audio systems.

The physical characteristics of such a sound field [5, 6] and their effects on human perception [2, 3] are not well understood. That is, current objective metrics fail to reliably characterize the physical properties of these sound fields in a robust and perceptually relevant way [7–11]. As a consequence, automotive audio manufacturers rely heavily on the human perception, as the instrument to characterize, evaluate, and optimize the sound quality of car audio systems. Typically, an iterative process is followed, where alterations on the audio system are perceptually evaluated, targeting the most pleasing aural experience [2, 12–14]. A number of studies investigated human perception in automotive audio by primarily focusing on comparative evaluations of (1) with–in audio system comparisons, such as preference on equalization [15, 16], DSP algorithms [13], and perceptual codecs [2], as well as (2) in–between audio systems comparisons and market benchmarking purposes [14, 17–19].

To the authors' best knowledge, there is no published literature on the perceptual effects of the acoustic transmission medium, the car cabin itself. Understanding the salient factors affected by the cabin's acoustics could aid the development of perceptually relevant models and metrics for assessing automotive audio. Moreover, it would depict the underlying relationship between physical and perceptual qualities of the car audio systems, enabling a more efficient optimization of the in–car aural experience.

In a recent study [20], it was shown that current perceptual evaluation protocols within automotive audio may not be able to faithfully capture the characteristics of cabin acoustics and a new assessment methodology was proposed. Here, this methodology is applied in the context of car cabin acoustics, where several physical modifications of a cabin's interior have been perceptually evaluated by expert assessors.

The aims of this study are: (1) to investigate the influence of acoustical properties of car cabins on the perceived qualities of the reproduced sound; (2) to identify the underlying relationships between physical and perceptual properties within car cabins; and, (3) to establish and further validate the applied experimental framework [20] followed, for assessing the acoustical properties of sound fields within automotive audio.

In Section II, the rationale behind the study is discussed and the experimental methodology is described. The data analysis is then presented in Sec. III, followed by the results and conclusions in Sec. IV and V, respectively.

2 Method

The experimental methods followed in this study include novel approaches in the acquisition and presentation of the captured sound fields as well as in the evaluation processes. This approach enabled the perceptual assessment of car cabin acoustics in laboratory settings. It further allowed human assessors to identify individually elicited perceptual attributes, which characterized both the spectral and the spatial properties of the sound fields under investigation; a serious limitation of previous studies [13, 16, 17, 20].

Spatial Decomposition Method (SDM) [21] is employed for recording and reproducing the sound fields to human assessors. As an alternative technique to binaural rendering [3, 16], this method eliminates several shortcomings related to binaural audio schemes [22], such as the lack of *externalization* and the subsequent difficulty in assessing spatial acoustics [13]. SDM has been successfully applied in the assessment of perceptual qualities of concert halls [23, 24], as well as in evaluating small-sized spaces, e.g., studio control rooms [25]. The applicability of SDM in automotive environments has been recently investigated and a recommendation was proposed [26].

Identifying the perceptual constructs underlying the cabin's physical properties would require a protocol where novel and uncommon aural experiences could be characterized and evaluated. This could be accomplished with descriptive *Sensory Analysis* (SA) [27] techniques. However, common SA procedures are time–consuming, laborious, and expensive, as they require product– and panel– specific training over multiple sessions [27, 43]. This is a major limitation in the time–restricted automotive industry. This paper applies a *rapid sensory analysis* method, i.e., the *Flash Profile* (FP) [28], and assesses its applicability within the automotive environment.

Several practical limitations exist in automotive audio assessment that FP seem to overcome. FP limits the required evaluation time by omitting the familiarization, the panel training, and the consensus vocabulary phases. Moreover, FP does not require product–specific training, compared to the traditional descriptive SA methods. This allows the use of assessors with general sensory expertise, requiring only 4–5 expert assessors for a statistically stable outcome [28, 29]. Nevertheless, FP is the closest rapid method to conventional descriptive SA [28] and it allows the quantitative description of stimuli, by statistically merging the quantitative and qualitative data in a common factorial space.

2.1 Experimental Design

The experimental design followed FP [30, 31] principles, adapted to assess audio material [20]. FP includes two experimental phases in a single session. First, each assessor is required to develop own set of perceptual descriptors during an attribute *elicitation* phase. Later, an attribute *ranking* phase is conducted where assessors comparatively quantify all stimuli simultaneously, by means of ranks, for each of the elicited attributes.

Two *Independent Variables* (IV) were included in the experiment. The acoustical *Condition* (twelve levels) combined with *Program* (three levels), resulting to a total of thirty–six stimuli. The ranking scores of each stimulus on the elicited attributes formed the quantitative *Dependent Variables* (DV).

2.2 Assessors

Four expert assessors participated in the experiment as volunteers. The assessors had 10–15 years of experience (Mean = 12, *s.d* ± 2.15) in critical listening, acoustical development, and sound tuning of premium automotive audio systems. As part of their profession, assessors have been trained to use their senses in critically evaluating the qualities of audio signals. They have all participated in numerous listening experiments and they were familiar with common SA procedures. All four assessors were male, aged 29–45 years old (Mean = 38). Their hearing sensitivity was confirmed to lie above 20dBHL between 125–8k Hz by standard hearing threshold procedures [32].

2.3 Materials & Apparatus

In a series of previous studies the apparatus has been described in detail including in–car acquisition of *Impulse Responses* (IRs), spatial analysis and synthesis of a car audio system using SDM [26], as well as the design and implementation of the experimental setup and the methodology [20] followed here. A brief description of these topics is given below.

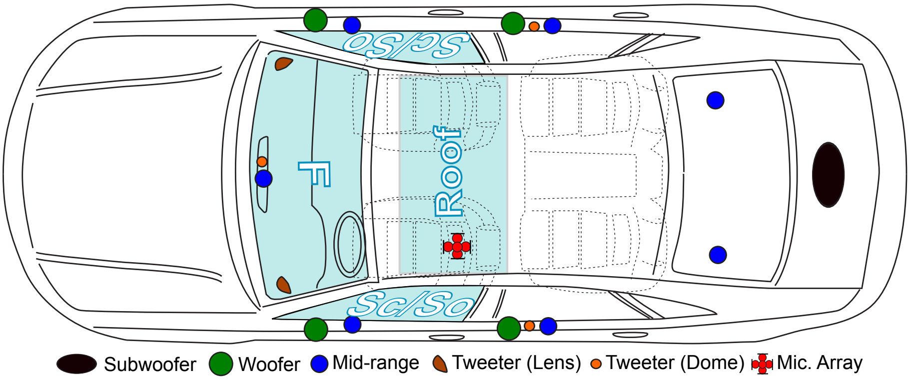


Fig. C.1: In–car audio system used in the measurements. The position of the microphone array is shown at the driver's position. Shaded areas, labeled as *Sc/So/F/Roof*, indicate the surfaces modified during the measurements (see descriptions in Table C.1).

In–situ Car Measurements

In order to capture the acoustical characteristics of the car cabin, in–situ measurements were performed in a sedan–type car (Audi – A8 Typ.4E), equipped with a premium audio system. The system comprised of 17 independent transducers driven by an automotive digital amplifier. The system is shown in Fig. C.1.

Individual spatial IRs were captured for each transducer of the system by a vector intensity probe (G.R.A.S 50VI–1), placed at the average seating position of the driver [33]. The measurements were conducted in a temperature and noise regulated garage ($V = 206m^3$, $RT_{30} < 0.2s$ at 125–8k Hz). These measurements are further referred to as *Vehicle Impulse Responses* (VIRs).

The signal path was set in such a way that the acquired VIRs included DSP processing (sound tuning), designed by a tonmeister. That included spectral and level balancing of the system, delays, and individual tuning of the speaker's magnitude responses in order to achieve a perceptually pleasing reproduction. This signal flow ensured that the experimental apparatus represented the performance of a typical premium automotive audio system.

Acoustical Conditions

Table C.1: Acoustical modifications used in the experiment. Condition *Ref* serves as the reference, representing a typical production car, equipped with premium audio system and no acoustical modifications. A dash indicates no alteration from car's reference settings.

Condition	Side Windows	Windshield	Ceiling	Cabin	DSP
Ref	–	–	–	–	–
EQ1	–	–	–	–	Alternative
EQ0	–	–	–	–	Disabled
Sc	Absorptive	–	–	–	–
So	Open	–	–	–	–
F	–	Absorptive	–	–	–
FSc	Absorptive	Absorptive	–	–	–
FSo	Open	Absorptive	–	–	–
Abs	–	–	–	Absorptive	–
AbsF	–	Absorptive	–	Absorptive	–
AbsSc	Absorptive	–	–	Absorptive	–
Roof	–	–	Reflective	–	–

The interior of the car cabin was systematically modified so that a representative range of possible acoustical fields was captured. The main compartments of the car's interior were altered in such a way that the first arriving reflections of the cabin's sound field were affected, as well as the later reflections, and combinations of both. The acoustical measurements obtained formed the experimental conditions as summarized in Table C.1.

Seen from the driver's seat perspective, the major reflection points in this car have been identified on the glass surfaces of the cabin, namely the side door windows and the windshield [20, 26]. In order to alter these reflections, 0.04 m thick Basotect foam [34] was used to fully cover the glass surfaces for condition *Sc*, as well as condition *F* and their combinations, conditions *FSc, FSo, AbsF* and *AbsSc*.

Although common room acoustics metrics, such as Reverberation Time, cannot be generalized in car cabins [5], it is common to observe decay times of 80 ms at mid-frequencies from measured IRs. To investigate the effects of the decay time of a car cabin, e.g. due to human occupancy or highly absorptive interior, another acoustical condition was included. A collection of absorptive materials was added to the cabin, including a 3.4 m×0.04 m rolled Acoustilux, with radius of 0.35 m, placed at the rear seats, highly absorptive fibre textiles placed at the interior's floor, and four pylons of Polyurethane foam sized 0.15×0.25×1.2 m on the front seats. Attention was given so that the direct acoustical paths between the sources and the receivers were not obstructed by the added materials.

Recently, car manufacturers have incorporated glass roofs instead of the conventional textile upholstery. In order to accommodate the effects of such scenario, condition *Roof* was included, where a unified glass tile (1.0×0.6×0.05 m) was attached to the ceiling of the cabin, positioned symmetrically above the front seats. During this condition, the absorptive nature of cabin's ceiling was altered to exhibit strong reflective characteristics. The topology and details of the above alterations are given in Fig. C.1.

The reference condition, indicated as *Ref* (Table C.1), refers to the captured sound field where the car cabin and DSP processing were unmodified, as the production automotive audio system, and it is further used as the baseline. Two additional conditions were included in the experiment, where only the DSP processing of the audio system was modified; the cabin's properties were kept at the reference settings. First, an alternative DSP processing preset was included, referred to as *EQ1*, where the door–woofers output was reduced −3 dB and the balance between the front left–center–right transducers was altered, aiming to increase the spatial width [35] compared to *Ref.* In addition, a condition where the system's DSP processing was disabled altogether, is referred to as *EQ0.* These two conditions were integrated in the experiment to assess the perceptual effects of sound tuning, compared to the physical alterations of the cabin's interior. Moreover, they could form the experimental anchors, as the physical alterations imposed on the system are known to elicit certain perceptual differences to experienced sound designers. In this way the validity of the method and the subsequent experimental results could be verified.

Reproduction System

In order to recreate the captured sound fields in the laboratory, a suitable reproduction system is required. For this study a 40.3 spherical loudspeaker array, depicted in Fig. C.2, was designed. The design of the loudspeaker array was based on a spatiotemporal analysis [36] of the aforementioned VIRs (Sec.II(C)), including additional measurements of twenty different types of cars and audio systems [20]. This analysis was essential to ensure that both the direct sound from car's transducers, as well as the subsequent reflections, were optimally reproduced in the laboratory. To limit the influence of the experimental room to the investigated sound field, the reproduction system was installed in an anechoic chamber[a].

Signals

The captured VIRs were processed with SDM [26]. The SDM is a spatial analysis and synthesis scheme where the sound field is decomposed in terms of pressure, direction and time, and encoded into a spatiotemporal domain [36]. The SDM–encoded signals are then divided into individual IRs, which are then used for synthesizing the sound field using a finite loudspeaker grid, by means of convolution with audio material.

Here, three audio excerpts were chosen based on the results of two pilot studies. The excerpts used were: (1) *Armin van Buuren feat. Ana Criado – I'll Listen (2012) | 0:15–0:30*, (2) *Melody Gardot – She don't know (Currency*

Of Man, 2015) | 2:01–2:16, (3) *Female Speech English (EBU SQAM – 2008) | 0:00–0:15* [37]. The sound excerpts formed the three levels of program, and are further referred to as dance, jazz, and speech, respectively. These signals were loudness–matched before convolution at 15 dB LUFS and perceptually validated by an expert listener in–situ. During the experiment the reference reproduction level was set to $75\mathrm{dB}_{\mathrm{LAeq(15s)}}$ at the listening position.

2.4 Procedure

First, the assessors were briefed about the experimental procedure and the principles behind FP protocol. As part of their introduction, a custom MAX/MSP interface [20] was presented, and the assessors performed a training session where no sound was provided. They were then guided inside the testing facility. The experiment was conducted in dark conditions and controls were imposed so that the assessors were unable to see the experimental apparatus until they completed the experiment.

The experimental process was controlled by the assessor over a self–paced and self–controlled software on a touch screen. The assessor was aware that there were no time limitations to complete the tasks. Short breaks were allowed and regularly recommended to avoid possible listening fatigue.

2.5 Attribute Elicitation

During the attribute elicitation phase, each assessor was asked to provide as many discriminant attributes as needed, to fully capture the perceived differences between the available stimuli. Emphasis was given as to provide precise, singular, non–redundant and low–level terms, that one could rate on a scale between a *High* and a *Low* intensity. It was also recommended to avoid hedonic and affective expressions relating to preference or acceptance [27]. Within the interface one could define the extreme intensity anchors of each attribute. For example, for the attribute ‘Loudness’, the assessor could define its scale anchors as ‘quiet’ and ‘too loud’, respectively.

During the elicitation phase all twelve conditions (Table C.1) were presented simultaneously on the screen labeled as A–L, as required by FP guidelines [28]. The order of the stimuli was kept constant within each session and randomized between assessors. The software provided the option to change program whilst listening to the same condition, so that perceptual differences between specific conditions could be explored over a variety of programs. Before completing the task, participants verified that their attribute list described the main perceptual differences between all thirty–six stimuli (3 programs × 12 conditions). At the end of the elicitation phase, an interview was conducted where the assessor provided short definitions for the elicited attributes.

2.6 Attribute Ranking

The second phase required the assessor to comparatively rank the experimental stimuli for each of the individually elicited attributes. At this stage, the evaluation followed a block design. The number of blocks was based on the number of the attributes given by that assessor. Each attribute was evaluated in three sequential trials, one for each program level. At each trial the stimuli were randomly assigned to 12 buttons labeled as A–L. The presentation order of the program levels and blocks was randomized, as required by standard audio evaluation procedures [38].

3 Data Analysis

The collected data included thirty–seven individually elicited attributes and their corresponding ranks for each of the presented stimuli, as shown in Fig. C.3. Several multivariate techniques could be followed to analyze such a dataset, e.g., *General Procrustes Analysis* (GPA) [39] and *Multiple Factor Analysis* (MFA) [40], both providing similar group–average patterns [29]. The mathematical transformations of MFA provide a number of complementary information, allowing the analysis of qualitative and quantitative data in a common latent space [41, 42].

In this study, the analysis is based on MFA, aiming to devise a common consensus space across assessors, while identifying the most important components, observations, and attributes [42, 45]. MFA studies the relations between several predetermined groups of attributes and it could be viewed as a consensus *Principle Component Analysis* (PCA), built on a set of equally–weighted principal components. MFA performs PCA on the attributes of each assessor separately, which are then normalized[b] to balance the influence of each group on the computation of the consensus space. The PCA data is then merged into a global matrix where a final PCA is performed, estimating the consensus solution across all assessors.

The outcome of such analysis is the positioning of stimuli on a consensus space. Similarly to a PCA, the interpretation of a stimulus position is based on its calculated coordinates on each dimension, the *factor scores*, and the corresponding variables explaining these dimensions, referred to as *variable loadings*. The inter–stimuli relationships are based on the relative distances between the stimuli's coordinates on the consensus *factor map*. The rationale behind this sensory profiling could be explained by projecting the variable loadings on the consensus space, creating what is known as the *variable map*. The advantage of analyzing FP data using MFA is the ability to jointly interpret these two quantities on a common factorial space. This approach enables the researcher to identify the underlying perceptual constructs of the stimuli profiles based on the structure within the data. The statistical analyses described in this section

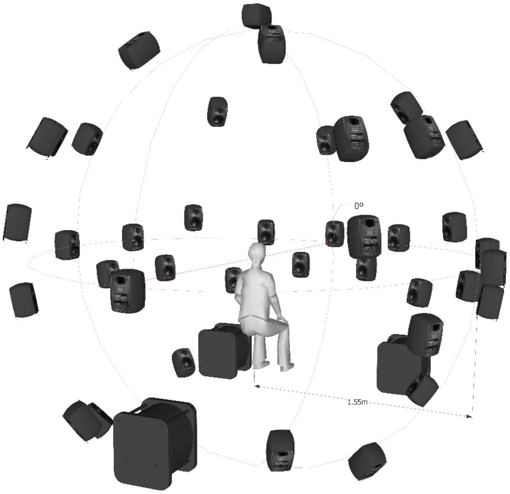


Fig. C.2: Reproduction system comprising of forty full–range loudspeakers and three subwoofers [20].

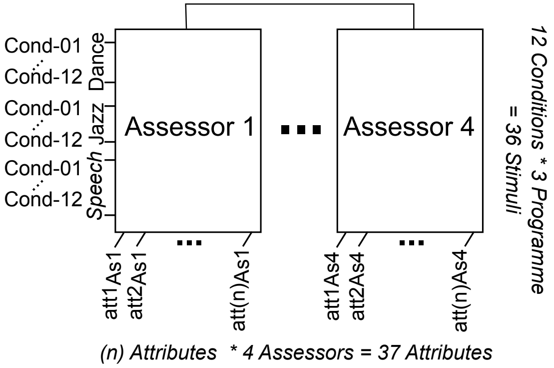


Fig. C.3: Data structure used for the MFA analysis, comprised of the observations of four assessors ($As(1-4)$) via a total of thirty–seven individually elicited attributes ($att(n)$), for the thirty–six stimuli.

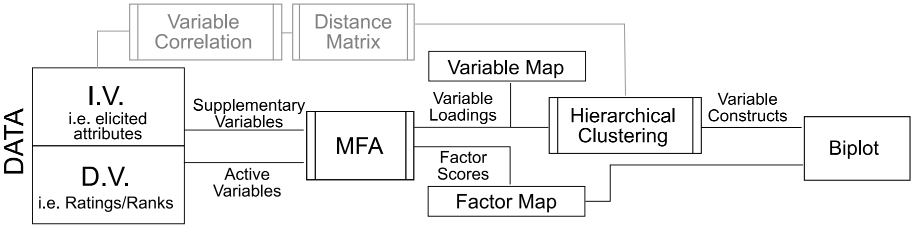


Fig. C.4: Schematic representation of the data analysis. Active variables refer to the data points used for the calculation of the latent dimensions using MFA. Supplementary variables do not contribute in the calculation of the dimensions but could be included for further statistical analysis; i.e. their correlation to the latent dimensions is visualized by projecting them into the MFA solution as vectors. The alternative clustering process followed for cross–validation of clusters hierarchy is shown in grey.

are summarized in Fig. C.4.

3.1 Ordination with Multiple Factor Analysis

MFA was performed on the collected observations (Fig. C.3) using *FactoMineR* package [46]. To reduce any scaling effects [47] between assessors, the raw data were centered, by subtracting the mean values of each column (attribute), and normalized, by dividing the centered data of each column by its root–mean–square. The analysis shows that almost 54% of the variance is explained by the first two principal components, and the remaining components seem to provide little contribution to the explained variance as shown in Table C.2. Figure C.5 shows the positions of the stimuli on the first plane, as a factor map. At this initial screening it can be seen that the stimuli are well separated in the first two common dimensions. *EQ0* and *AbsF* hold the extreme positions on dimension 1, whilst *Roof* contrasts those two conditions, on dimension 2. *Ref*, *EQ1* and *Roof* are positioned relatively close to each other in both dimensions, as expected, due to their subtle audible differences. Moreover, it can be seen that the more absorption added in the cabin, the more negative the dimension 1 becomes for these stimuli. This can be observed by contrasting the baseline's dimension 1 coordinates (*Ref*) to the condition where absorption is added in the cabin (*Abs*). Adding absorption material on the side windows (*AbsSc*) and the windshield (*AbsF*) continues to have a negative effect on dimension 1.

3.2 Influence of Program & Acoustical Conditions

A common way to identify significant differences within the stimuli–set in the latent MFA space follows the calculation of 95% *Confidence Ellipses* (C.E.) [48], an analogous metric to *Confidence Intervals*. The C.E. of condition levels are depicted in Fig. C.5, indicating good separation between most condition levels,

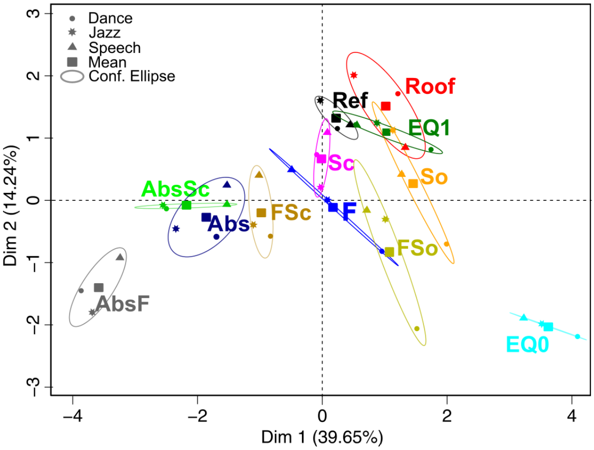


Fig. C.5: MFA consensus space depicting the coordinates of the twelve conditions included in the experiment on the first two principal components. The points indicate the factor scores of each of the thirty–six stimuli, signified by their program level and colored by the condition level (see Table C.1). The 95 % Confidence Ellipses depict the significant differences between conditions.

as seen on the first two dimensions. These observations indicate that the panel of four expert assessors provided sensory ratings that are significantly different between different conditions.

Moreover, it is noted that most of the thirty–six stimuli are clustered together, in groups of three, following their corresponding condition level. No systematic bias can be seen for a specific program excerpt, or extreme values that would indicate program-dependence of the acoustical conditions. This indicates that stimuli are ordered similarly even when different program excerpts were used. Further analysis verified that the program have no significant effect on the perceived differences within the various conditions ($R^2_{\text{Dim1}} = 0.001$, $R^2_{\text{Dim2}} = 0.04$, $p = n.s.$) [46].

3.3 Generalizing Results – Averaging

In order to achieve a holistic understanding of the data and focus on the IV of interest, the acoustical conditions, a MFA was performed on the averaged data across program. This approach addresses the relatively low explained variance of the first two dimensions (54%) of the previous analysis by accounting for the noise within the data at lower dimensions. Moreover, since the program was found not to be a significant factor, the relative positions of the stimuli would be preserved.

The variances explained by the first five dimensions of the MFA analysis on averaged data are summarized in Table C.3. The first two dimensions explain

Table C.2: The first five Principle Components of the MFA analysis, based on the analysis of the normalized and centered data.

P. Comp.	Eigenvalue	% of Var.	Cumul.% Var.
1	3.53	39.65	39.65
2	1.26	14.24	53.89
3	0.53	6.01	59.90
4	0.45	5.15	65.05
5	0.34	3.91	68.97

Table C.3: The first five Principle Components of the MFA analysis based on the averaged data across program.

P. Comp.	Eigenvalue	% of Var.	Cumul.% Var.
1	3.78	56.85	56.85
2	1.23	18.61	75.47
3	0.42	6.36	81.84
4	0.36	5.53	87.38
5	0.24	3.68	91.06

75.47% of the variance and there is minimal contribution by the remaining individual dimensions (< 7%). Figure C.6 depicts the factor scores based on the analysis of the averaged data. As expected, the relative positions between the conditions are very similar to the ones in Fig. C.5, yet explained by notably higher variance.

The variable map, shown in Fig. C.7, depicts the attributes of each assessor projected to the MFA plane. A high number of attributes is well represented on the first two dimensions, making the graphical interpretation a difficult task. Reducing the number of attributes would enable a better interpretation of the results. That is, classifying the assessors' own attributes into collective categories and in consequence into the common underlying perceptual constructs.

This could be done semantically, based on the homologous terms and the definitions given by assessors. However, in FP each assessor uses own vocabulary, thus, an attribute given and scored by an assessor may not necessarily relate to a semantically–similar attribute given by another assessor [41]. A mathematically–based approach i.e. using the geometrical and the statistical properties of the data points [42], would reveal the true structure within the dataset. Combined with the definitions given by assessors the *internal validity* of the formed clusters could be assessed. That is, the extent on which the grouped variables measure and represent similar sensory constructs. Recently, such methods were successfully applied on individually elicited attributes of

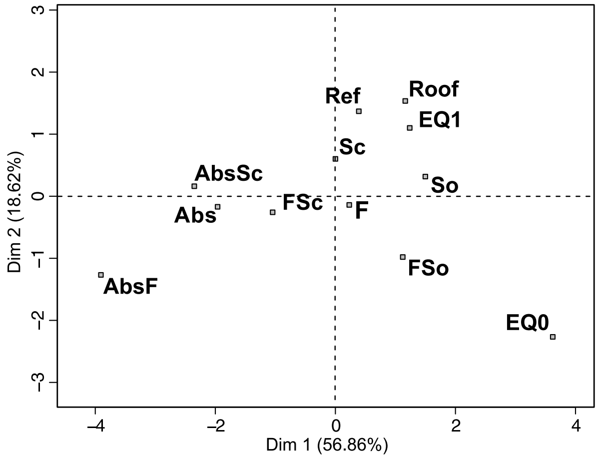


Fig. C.6: The resulting factor map of the MFA analysis using averaged data across program. The map depicts the position of stimuli on the panel's consensus space.

audio material, and allowed the identification of the common perceptual constructs across thirty–one [43] and twenty–three naive assessors [44].

3.4 Clustering of Elicited Attributes

The grouping of attributes was achieved using *Agglomerative Hierarchical Clustering* (AHC), based on the Euclidean distances of the MFA coordinates of each attribute, in conjunction with Ward's criterion [43, 49]. As clustering is blind to the importance of each attribute to each dimension, thus, susceptible to noise, the attributes included in the analysis were pre–selected based on the correlation of the attribute to any the first two principal components (r>0.65). This noise reduction process accounts for these limitations of AHC and the clustering process provides equal hierarchical weights between the well-correlated variables only.

Two main clusters can be identified in the resulting dendrogram in Fig. C.8. The first cluster is formed by two subcategories, one described by attributes related to *Bass*, and one related to the spatial *Image focus*[d]. The second cluster splits into four subcategories and includes attributes related to *Ambience, Width & Envelopment, Transparency*, and *Brightness.* It is noted that the attributes clustered well together semantically, especially for the clusters related to *Bass, Brightness, Image focus* even if no consensus vocabulary or panel training was included in the procedure. The attributes related to *Width* and *Envelopment* fall into the same cluster. This comes in agreement with previous studies [35], where assessors used these attributes interchangeably as they both contributed to the perceptual construct of spaciousness [50]. Yet, it could also

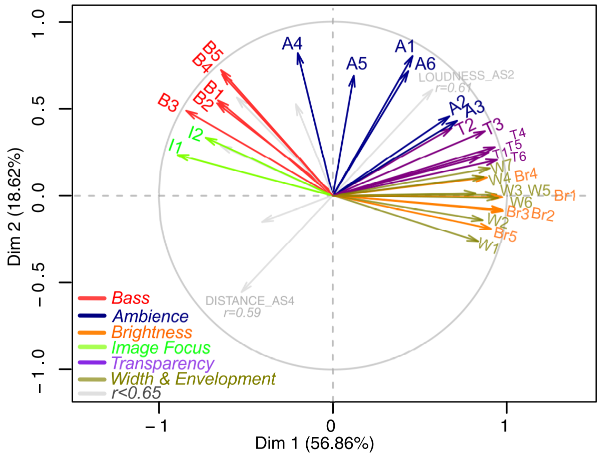


Fig. C.7: Variable map from MFA analysis if the program-averaged data, depicting the projections of the individual attributes as vectors, on the first consensus plane. The length of the vector indicates the correlation to the factorial solution. The vectors' colors indicate their cluster group, as calculated in Sec. III(D). The attributes' labels referred to the convention used in Fig. C.8. Grey vectors indicate the excluded attributes and include the label and assessor's number.

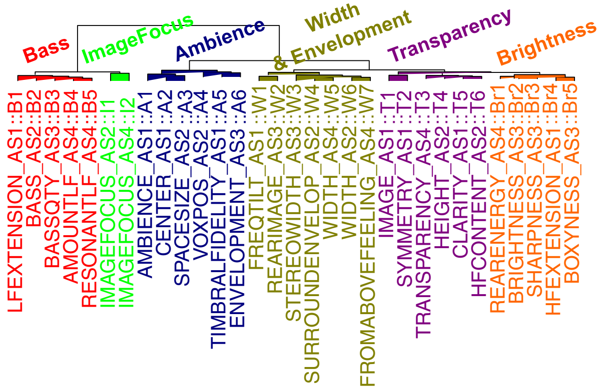


Fig. C.8: Dendrogram of the individually elicited attributes. The clustering processes was based on the MFA coordinates of the averaged dataset. The first eight dimensions were used in the clustering, based on euclidean distances and ward's criterion. AS$(1-4)$ denotes the assessor's number.

indicate that the stimuli used in the experiment failed to excite these constructs separately.

These six clusters observed here are thought to encompass the perceptual constructs underlying the stimuli set in this investigation. Although the individual attributes may differ within a cluster, e.g. 'Image_AS1' has been grouped under ‘Transparency’ , the clustering algorithm identified a perceptual equivalence across the grouped attributes. That is, the assessors rated similarly the stimuli for these attributes, even if they are not semantically related; a common observation in free-elicitation experiments [41]. Here, the clusters were labeled following the definitions given by the assessors during evaluation, in combination to previous studies on spatial acoustics [35, 51] and sound reproduction [52] to maintain consistency across studies and illustrates the author’s best understanding.

It should be noted that the input to the clustering algorithm used here, included the coordinates of the attributes on all dimensions given by the MFA analysis. This allows to directly project clusters on the latent MFA space, seen in Fig. C.6. Yet, AHC might produce hierarchies for objects that are not hierarchically interrelated [53]. To validate the clustering process an additional AHC was performed. Using the raw data, the correlation matrix of the attributes was used as the input of the AHC in the form of a distance matrix [42]. This clustering revealed similar results to the original clustering, confirming the validity of the dendrogram depicted in Fig. C.8 and the perceptual constructs identified.

4 Results

The interpretation of the data can be achieved by graphically combining the results of the statistical analysis described above, in the form of a Biplot. That is, merging the consensus factor scores of the MFA (Fig. C.6) and the perceptual constructs identified by AHC. To achieve this, the MFA coordinates of the individually elicited attributes are averaged per cluster, and then projected into the MFA factorial space (Sec. III(C)). This process allows the efficient visualization of the results, by simultaneously presenting the major quantitative and qualitative observations. Figure C.9 depicts the summarized results of this paper, combining the factor scores and the identified perceptual constructs. The perceptual constructs are realized as directional vectors, providing a tangible explanation for the variance within each dimension.

As the conditions are a combination of several factors including modification of: (1) Front Side Windows, (2) Windshield, (3) Roof, (4) Cabin Absorption, and their combinations, the best interpretation of this graph is achieved by analyzing comparatively conditions where single changes occurred.

4.1 Validation – Effect of Equalization

First, by focusing on the conditions where only the DSP settings were modified, i.e. *EQ0, EQ1*, one could verify whether the identified factor space and the related perceptual constructs come in agreement with our expectations and

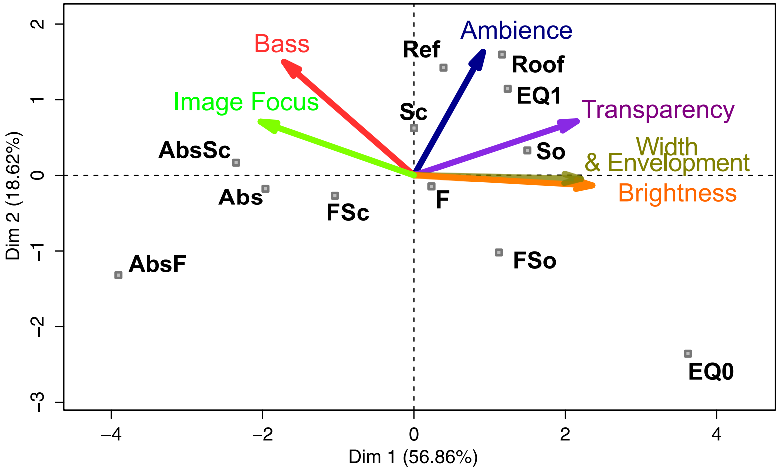


Fig. C.9: Biplot depicting the perceptual constructs and the stimuli factor scores. The identified constructs are projected in the factorial space, by calculating the average coordinates of each cluster's attributes.

current knowledge.

The DSP settings of *EQ0* included a substantial reduction at low frequencies (−15 dB) compared to the reference DSP settings (*Ref*), which explains its projection to the basis vector being directly opposite to the perceptual construct of Bass, as seen in Fig. C.9. This denotes that assessors identified a decreased low frequency content when listening to *EQ0*. The minor differences in the low frequencies of *EQ1* compared to *Ref* have also been perceived by the assessors at the appropriate intensity level, indicating only a slight increase of Bass for *EQ1* compared to *Ref*. The close positioning of the *EQ1* and *Ref* supports that audible differences were subtle, as noted. In contrast the extreme position of *EQ0* in Fig. C.9 indicates perceptually strong differences compared to the other stimuli.

Moreover, *EQ1* included slight alterations on the spatial and spectral balance of the front channels. The constructs of Width & Envelopment and Brightness indicate slightly higher intensities of the *EQ1* compared to *Ref*, on the expense of reduced Image focus. The position of *EQ0* indicates high values against these perceptual constructs on dimension 1 which follows the expected results, as the sound tuning of automotive system is based on the optimization of such constructs.

These observations suggest that the evaluation method successfully captured the perceptual differences across stimuli, depicting the underlying perceptual factors and the relative intensities in an expected way. That is, the ability of the experimental apparatus to facilitate the perceptual differences across the measured sound fields and its capacity to identify and signify these differences by employing the statistical procedures described above.

4.2 Effect of Roof

A glass added on the ceiling of the cabin would increase the energy coming from above the listener, especially at higher frequencies. The number of the very first arriving reflections may also increase, affecting the echo density of the field as well as the perceived interaural differences. These physical alterations have been linked to the perceptual attributes of apparent source width, envelopment, and spaciousness [35].

Based on Fig. C.9, *Roof* is closely positioned to *Ref*, indicating slight perceived differences. The main differences can be seen on dimension 1. This indicates that adding a glass to the ceiling increases the perceived Width & Envelopment, Brightness, and Transparency. Minor increase is noticed on the perceived Ambience. Image focus is however decreased, which could relate to comb filtering, due to the added early reflections of the ceiling [54]. That is a spectral interference of coherent signals which may perceived as spectral and spatial alterations of the originally emitted signal. Interestingly, *Roof* and *EQ1* seem to hold equal positions in dimension 1, indicating their perceptual similarities.

4.3 Effect of Absorption

Acoustic damping materials were added in the cabin, aiming to decrease the decay time of the cabin's sound field and consequently the perception of the apparent room size [35] and reverberance [44]. The position of *Abs* compared to *Ref* indicates a much less Ambient field, less Wide & Enveloping, yet, a more focused sound Image. This observation may relate to the decreased number of strong and dense reflections from multiple directions. Perceived Brightness is also decreased, as expected, due to the highly absorbing materials used, at this frequency range. This spectral imbalance of the system could be also observed by the identified increased Bass content.

These observations are also supported by contrasting the position of conditions *F* and *AbsF*, as well as the *Sc* and *AbsSc*; all indicating perceptual equivalence of conditions when the absorption in the cabin increases.

4.4 Effect of Front Side Windows

Two conditions aimed to investigate the perceptual effects imposed by the reflections of the front side windows. In condition *So* the windows were open, so that the reflections originating from the glass surface were eliminated, and the cabin was an acoustically open–cavity. In a second alteration, *Sc*, the door's glass surface was covered with absorptive material, aiming to decrease the subsequent reflected energy.

The two conditions, *So* and *Sc*, revealed dissimilar sensory profiles as seen in Fig. C.9, even if the same surface was altered. When absorption was added (*Sc*), the perceived Bass was highly similar as to *Ref.* In contrast the *So* condition, where the windows were open, the amount of low–frequency energy in the cabin was perceived as reduced. That is an expected result, as the absorption material used in the experiment was only affecting high frequencies; opening the windows should also affect the modal behavior of the cabin. Similarly, an increase is apparent in the perceived Width & Envelopment at *So*, compared to when windows were covered with absorption material, as in *Sc*.

These trends can also be seen by comparing the conditions *F–FSo*, where the relative difference between the two was the front side windows state. The relative distances and projections of this pair indicate their perceptual similarities to *Ref–So*. This may indicate that the perceptual effects of opening the side window are independent of the windshield properties.

Side Windows – Absorptive

The effect of increasing the absorption of the side windows seems to decrease the perceived Ambience, and the Bass content at a lower degree. No major alterations can be seen on the perceived Width & Envelopment and Brightness. This would be an unexpected observation for room acoustics, as the side reflections are known to affect these perceptual constructs [55, 56]. One could hypothesize that such result may indicate a different auditory processing scheme [57] when exposed to car's sound fields; due to the highly dense early reflections that arrive within a few milliseconds, in contrast to the distinct and sparse reflections in typical rooms [55]. Yet, it should be noted that the car audio system was equipped with acoustic lens technology [58] at the front tweeters, where the dispersion area of the high frequencies is optimized. Therefore such reflections could be limited in this experimental setup. Thus, the expected effect on spatial width may not have been perceived in this investigation.

Side Windows – Open

To investigate the effect of the side windows in a different way, the glass surface of the front doors was removed. Based on the positions of *Ref* and *So* in Fig. C.9, it can be seen that the perceived Ambience is less apparent when windows are open. The sound is also perceived slightly more Wide & Enveloping and Brighter, whereas the Image focus decreases. This condition should indeed affect the perceived Ambience, as the cabin was not a closed–cavity anymore. Moreover, perceived bass is affected, in agreement with previous findings of possible standing waves along that direction [26] and increased room gain in car cabins [59]. These findings are also supported by examining the factor scores of *F* and *FSo*, as their relative positions are highly similar to *Ref* and *So*.

4.5 Effect of Windshield

The perceptual effects of adding absorptive material on the windshield, referred to as *F*, are mainly apparent on the second dimension compared to reference condition. The reduced energy coming off the large glass surface opposite the driver seems to reveal a less Ambient and less Transparent sound field, yet, Width & Envelopment and Brightness are not affected.

However, comparing *Sc* and *FSc*, where the relative physical change between conditions was identical to *Ref* and *F*, perceptual effects are apparent also on dimension 1. This is an intriguing result which may indicate a strong relationship between the combined Front and Side reflections, on the perceived spatial properties, when the cabin is a closed-cavity.

5 Conclusions

The study employed a recently proposed evaluation methodology for automotive audio, to address the perceptual effects of car cabin acoustics. The experimental methodology included the *Spatial Decomposition Method* for the acquisition, analysis and presentation of the sound fields to human assessors, whilst a rapid sensory analysis protocol, *Flash Profile*, was adapted and used for audio material. The method provided individual vocabulary profiling from expert assessors, in a single experimental session of 1.5–3h in total.

The findings indicate the importance of the acoustical properties of a car cabin on perceived sound quality. It was shown that even slight alterations in the cabin, for example adding a reflective glass surface above the listener, have a notable impact on the perceived sound field. Moreover, the significance of reflections originating from the windshield was identified as in a previous study [60], as well as the influence of the side windows on the perceived sound, and a relationship between the two surfaces was also apparent. The optimization of the system by means of equalization and DSP processing seems to highly alter the aural experience, supporting the relevance of the industry's current sound tuning approaches. Finally, the identified effect of added absorption, even at extremely short decay times may reflect on the proposed influence of passenger occupancy [3] on the reproduced sound in cars. One could infer relations of these results to previous investigations that sought to identify the perceptual aspects of sound in enclosures, e.g., studies in concert halls [43, 44] and sound reproduction in small rooms [55]. A comprehensive literature review [35] suggested that the perceptual space characterizing performance spaces, sound reproduction in domestic rooms, and automotive audio seem to be heterogeneous. The current results support this notion. That is, similar trends could be observed but the interrelations of the perceptual constructs differ, and a direct comparison would be an inaccurate representation. As similar studies in the domain of automotive audio were not identified, the perceptual space cannot be contrasted directly to previous results. Here, specific findings were compared to related literature.

The investigation demonstrates the applicability of the Flash Profile in perceptual evaluation of automotive audio systems. Further, it allowed a statistically robust characterization of the stimuli-set based on multivariate analyses of both quantitative and qualitative data. The underlying perceptual constructs of the sound fields were identified, and projected against a data–driven factorial analysis. Two stimuli–anchors were used in the experiment, *EQ1* and *EQ0*, where their factorial position and perceptual interpretation comes in agreement with our expectations and empirical knowledge. This further validates the experimental design [20] and the subsequent data analysis.

Nevertheless, several challenges [20] should be addressed when FP is applied in audio. That is, the assessors should be carefully selected, based on their general sensory abilities and background [20], as the quality of the given descriptors is highly important [61]. One should note that *Flash Profile* is not intended to provide a robust attribute vocabulary. Here a number of steps were followed to improve this limitation of the FP protocol, e.g., by introducing a short interview where definitions were given, and by recruiting highly experienced and product–expert assessors. Moreover, stimuli–anchors were added in the experimental design and a careful statistical analysis was followed.

The use of different program types is necessary for audio evaluation [38]. Thus, the stimulus that one aims to evaluate, is the product of a program (i.e. speech) and an acoustical modification (i.e spectral alteration). This is not a parameter that *Flash Profile* and the associated statistical methods account for imposing practical and statistical challenges [20]. For example, FP requires all stimuli to be available to the assessor simultaneously. During attribute elicitation, this could be accommodated. However, during the ranking phase, a block design is followed as the acoustical conditions must be evaluated for each program material separately [38]. In consequence when analyzing the results, one should follow statistical procedures that allow a two-way interaction, between the program and the condition used. Here, the data were averaged across program levels before the final MFA analysis to overcome this limitation[e], as no significant difference between program levels was identified. Hierarchical clustering was then used to obtain the common perceptual constructs within the collected data and enable an interpretation of the results at the panel level. The two analyses were then merged in the form of a biplot. That is, the data–based solution of factor scores, based on the stimuli rankings, and the perceptual constructs identified via attribute clustering, indicating the direction of the explained variance.

The current results provided evidence that the proposed method allows the perceptual assessment of audio material within car cabins, and contributed to our knowledge of cabin acoustics. It depicted the importance of several surfaces of car's interior and the perceptual relationships to such changes. The investigation assessed a limited number of acoustic modifications in the cabin, aiming to explore cabin acoustics and assess the applicability of the method in the automotive environment. It is however, the first time that such an investigation is conducted in car cabins. Thus, further validation studies should be conducted. Further studies will improve our knowledge of car cabin acoustics and identify ways to compensate for the related sound degradation. Moreover, objective metrics such as spatiotemporal analysis [36] could supplement the perceptual data presented here, as shown previously [20, 26]. This would allow a better understanding of the acoustical fields and robust investigation, supported by both physical and perceptual metrics.

Future work includes the investigation of several cabin acoustics and systems as well as the application of the method in other acoustical environments, for example small–sized residential rooms and listening spaces.

Notes:

[a] The anechoic chamber has free inner dimensions of 5.0 × 4.5 × 4.0 m, and meets the requirements for anechoic performance down to 200 Hz. Below this point, a low frequency compensation has been applied, as detailed previously [20].

[b] PCA results are divided by the square root of the first eigenvalue, i.e. largest singular value.

[c] The explained variance of a variable to the factorial space is given by $R = \sqrt{R^2_{\text{Dimension1}} + R^2_{\text{Dimension2}}}$. Limiting the variable's correlation the dimension allows a factorial, i.e. exploratory, analysis whereas the total correlation to the factorial solution is controlled, yet, the subspace is not limited to two-dimensional inertia, i.e. highly correlated to only one dimension.

[d] *Image focus* was defined by both AS2 and AS4, as the extend at which the sound source appears to be at certain location. The scale anchors were labeled as muffled and focused for low and high intensities, respectively.

[e] Alternatively, or in the case of the averaging is not possible, *Hierarchical Multiple Factor Analysis* (HMFA) [62] techniques could be followed to facilitate similar analysis.

Acknowledgments

The authors would like to thank Morten Lydolf, Martin Møller, Martin Olsen, Claus Vestergaard Skipper, and their colleagues at Aalborg University for their support and helpful input. The research leading to these results has received funding from the European Union's Seventh Framework Program under grant agreement no. ITN-GA-2012-316969.

References

[1] F. Rumsey, "Automotive Audio", J. Audio Eng. Soc. **60**(12), pp. 1070–1074 (2012).

[2] S. Bech, M.A. Gulbol, G. Martin, J. Ghani, and W. Ellermeier, "A listening test system for automotive audio – Part 2- Initial verification", in *the 118th Conv. Audio Eng Soc.*, May, 28–31, 2005, Barcelona, Spain, (Audio Engineering Society, New York), Prepr. No 6369.

[3] F. Christensen, G. Martin, M. Lydolf, M. Pauli, S. Woo-Keun, and P. Benjamin, "A Listening Test System for Automotive Audio - Part 1 : System Description", in *the 118th Conv. Audio Eng Soc.*, May, 28–31, 2005, Barcelona, Spain, (Audio Engineering Society, New York), Prepr. No 6358.

[4] E. Granier, M. Kleiner, B. Dalenbäck, and P. Svensson, "Experimental Auralization of Car Audio Installations", J. Audio Eng. Soc. **44**(10), pp. 835–849 (1996).

[5] C. Choi, L.-h. Kim, S. Doo, Y. Oh, and K.M. Sung, "Assessment of Sound Field in a Car", in *the 113rd Conv. Audio Eng. Soc.* Los Angeles, USA,Oct. 5-6, 2002, (Audio Engineering Society, New York), Prepr. No 5701.

[6] M. Kleiner and C. Lindgren, "Objective Characterization of Audio Sound Fields in Automotive Spaces", in *Proc. 15th Int. Conf. Audio Eng. Soc.*, Copenhagen, Denmark, October, 31-2, 1998, (Audio Engineering Society, New York).

[7] M. Straub and D. de Vries, "Application of Multichannel Impulse Response Measurement to Automotive Audio", in *the 125th Conv. Audio Eng. Soc.*, San Fransisco, USA ,October, 3-5, 2008, (Audio Engineering Society, New York).

[8] M. Strauss, J. Nowak, and D. de Vries, "Approach to sound field analysis and simulation inside a car cabin", in *Proc. 36th Int. Conf. Audio Eng. Soc. Automot. Audio*, June 2-4, 2009, Dearborn, Michigan, USA, (Audio Engineering Society, New York).

[9] A. Celestinos, O. Martin, M. Bo Møller, and M. Lydolf, "Car interior simulation model for low frequencies using the finite difference time domain method", in *Proc. 48th Int. Conf. Audio Eng. Soc.*,Munich, Germany, September, 21-23, 2012, (Audio Engineering Society, New York).

[10] J. Nowak and M. Strauß, “Sound Field Reproduction Analysis in a Car Cabin Based on Microphone Array Measurements”, in *Proc. 48th Int. Conf. Audio Eng. Soc. Automot. Audio*, Munich, Germany, September, 21-23, 2012, (Audio Engineering Society, New York).

[11] M. Binelli, A. Venturi, A. Amendola, and A. Farina, “Experimental analysis of spatial properties of the sound field inside a car employing a spherical microphone array”, in *the 130th Conv. Audio Eng. Soc.*, London, UK, May, 13-16, 2011, (Audio Engineering Society, New York), Prepr. No 8338.

[12] G. Martin and S. Bech, “Attribute Identification and Quantification in Automotive Audio - Part 1: Introduction to the Descriptive Analysis Technique”, in *the 118th Conv. Audio Eng Soc.*, May, 28–31, 2005, Barcelona, Spain, (Audio Engineering Society, New York), Prepr. No 6360.

[13] P. Hegarty, S. Choisel, and S. Bech, “A listening test system for automotive audio - Part 3: Comparison of attribute ratings made in a vehicle with those made using an auralization system”, in *the 123rd Conv. Audio Eng. Soc.* ,New York, USA, October, 5-8, 2007), (Audio Engineering Society, New York), Prepr. No 7224.

[14] S. E. Olive, “A New Reference Listening Room For Consumer, Professional and Automotive Audio Research”, in *the 126th Conv. Audio Eng. Soc.* ,Munich, Germany, May, 7-10, 2009, (Audio Engineering Society, New York), Prepr. No 7677.

[15] P. Dennis, “In-Vehicle Audio System Sound Quality Preference Study”, in *Proc. 139th Conv. Audio Eng. Soc.*, New York, USA, October, 29-1, 2015, (Audio Engineering Society, New York), Prepr. No 9393.

[16] S. E. Olive and T. Welti, “Validation of a Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems”, in *Proc. 36th Int. Conf. Audio Eng. Soc. Automot. Audio*, June 2-4, 2009, Dearborn, Michigan, USA, (Audio Engineering Society, New York).

[17] S. Squartini, P. Francesco, T. Romolo, N. Massimo, L. Walter, B. Ferruccio, C. Emanuele, and L. Ariano, “Evaluating different Vehicle Audio Environments through a novel Software-based System”, in *the 116th Conv. Audio Eng. Soc.*, May, 6-8, 2004, Berlin, Germany, (Audio Engineering Society, New York), Prepr. No 6083.

[18] D. C. Mikat, “Subjective Evaluations of Automotive Audio Systems”, in *the 101st Conv. Audio Eng. Soc.*, November, 8-11, 1996, Los Angeles, CA, USA, (Audio Engineering Society, New York), Prepr. No 4360.

[19] A. Farina and E. Ugolotti, “Subjective Comparison of Different Car Audio Systems by the Auralization Technique”, in *Proc. 103rd Conv. Audio Eng. Soc.*, September, 26-29, 1997, New York, USA, (Audio Engineering Society, New York), Prepr. No 4587.

[20] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S.H. Jensen, “A Rapid Sensory Analysis Method for Perceptual Assessment of Automotive Audio”, J. Audio Eng. Soc. **65**(1/2), (2017).

[21] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, “Spatial decomposition method for room impulse responses”, J. Audio Eng. Soc. **61**(1/2), (2013).

[22] S. Sun, Y. Shen, A. E. S. Member, and Z. Liu, “The Effects of Recording and Playback Methods in Virtual Listening Tests”, J. Audio Eng. Soc., **63**(7/8),pp. 570–582 (2015), http://dx.doi.org/10.17743/jaes.2015.0058 .

[23] H. Tahvanainen, J. Pätynen, and T. Lokki, “Analysis of the Seat-Dip Effect in Twelve European Concert Halls”, Acta Acust. united with Acust. **101**, pp. 731–742, (2015), http://dx.doi.org/10.3813/AAA.918869.

[24] A. Haapaniemi and T. Lokki, “Identifying concert halls from source presence vs room presence”, J. Acoust. Soc. Am. **135**(6), pp. 311–317 (2014).

[25] S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki, “Preferences of critical listening environments among sound engineers”, J. Audio Eng. Soc. **62**(5), pp. 300–314, (2014).

[26] S. Tervo, J. Pätynen, N. Kaplanis, M. Lydolf, S. Bech, and T. Lokki, “Spatial Analysis and Synthesis of Car Audio System and Car-Cabin Acoustics with a Compact Microphone Array”, J. Audio Eng. Soc **63**(11), pp. 914–925 (2015), http://dx.doi.org/10.17743/jaes.2015.0080.

[27] H. T. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices* (Springer: NewYork), (1999) pp. 1–471.

[28] J. Delarue, B. J. Lawlor, and M. Rogeaux, *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research* (Woodhead Publishing, Elsevier: Cambridge, UK), (2014), pp. 119-148.

[29] P. Varela and G. Ares, *Novel Techniques in Sensory Characterization and Consumer Profiling* (CRC Press: Boca Raton, USA) (2014), pp. 175–206.

[30] V. Dairou and J.-M. Sieffermann, “A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile”, J. Food Sci. **67**, pp. 826–834 (2002).

[31] J. Delarue and J. M. Sieffermann, “Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products”, Food Qual. Prefer. **15**, pp. 383–392 (2004).

[32] “International Organization for Standardization, ISO:8253-1 Audiometric test methods Part 1: Pure-tone air and bone conduction”, (2010).

[33] S. Parkin, G. M. Mackay, and A. Cooper, “How Drivers Sit in Cars”, Accid. Anal. Prev. **27**, pp. 777–783 (1995).

[34] BASF, “BASOTECT Foam”, http://product-finder.basf.com/group/corporate/product-finder/en/brand/BASOTECT, Accessed 10.01.2016.

[35] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, “Perception of Reverberation in Small Rooms : A Literature Study”, in *Proc. 55th Int. Conf. Audio Eng. Soc. Spat. Audio*,August, 3-5, 2014, Helsinki, Finland, (Audio Engineering Society, New York).

[36] J. Pätynen, S. Tervo, and T. Lokki, “Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses.”, J. Acoust. Soc. Am. **133**, pp. 842–57 (2013).

[37] European Broadcasting Union, “Tech 3253 - SQAM: Sound Quality Assessment Material recordings for subjective tests”, pp. 1–13, EBU: Geneva, (2008).

[38] S. Bech and N. Zacharov, *The Perceptual Audio Evaluation: Theory, Method and Application*, (Wiley: Chicester, England), (2006), pp. 1–442.

[39] J. Gower, “Generalized procrustes analysis”, Psychometrika, **40**(33), (1975),doi: 10.1007/BF02291478 .

[40] B. Escofier and J. Pagès, “Analyses factorielles simples et multiples: objectifs, méthodes et interprétation”, “Simple and multiple factorial analysis: objective, methods and interpretation” Dunod - Sciences Sup, (1998),.

[41] S. Lê and T. Worch, *Analyzing Sensory Data with R* (CRC Press: Boca Raton, USA) (2015), pp. 69–102 & 173–198.

[42] J. Pagès, *Multiple Factor Analysis by Example Using R*, volume 20 (CRC Press: Boca Raton, USA) (2014), pp. 78 & 189–208.

[43] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo, “Concert hall acoustics assessment with individually elicited attributes”, J. Acoust. Soc. Am. **130**(2), pp. 835–49 (2011).

[44] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, “Disentangling preference ratings of concert hall acoustics using subjective sensory profiles”, J. Acoust. Soc. Am. **132**(5), pp. 3148–3161 (2012).

[45] H. Abdi, L. J. Williams, and D. Valentin, “Multiple factor analysis: principal component analysis for multitable and multiblock data sets”, Wiley Interdiscip. Rev. Comput. Stat. **5**, pp. 149–179 (2013).

[46] S. Lê, J. Josse, and F. Husson, “FactoMineR : An R package for multivariate analysis”, J. Stat. Softw. **25**, pp. 1–18 (2008).

[47] P. B. Brockhoff, “Statistical testing of individual differences in sensory profiling”, Food Qual. Prefer. **14**, pp. 425–434 (2003).

[48] M. Cadoret and F. Husson, “Construction and evaluation of confidence ellipses applied at sensory data”, Food Qual. Prefer. **28**, pp. 106–115 (2013).

[49] A. Lucas, “amap: Another Multidimensional Analysis Package”, (2008), URL `https://cran.r-project.org/web/packages/amap/index.html, Rpackageversion0.8-4;Accessedon15.11.2015`.

[50] T. Okano, L. L. Beranek, and T. Hidaka, “Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls.”, J. Acoust. Soc. Am. **104**(1), pp. 255–65 (1998).

[51] A. Lindau, V. Erbes, S. Lepa, H.-j. Maempel, F. Brinkman, and S. Weinzierl, “A spatial audio quality inventory for virtual acoustic environments (SAQI)”, Acta Acust. united with Acust. **100**, pp. 984–994 (2014).

[52] T. H. Pedersen and N. Zacharov, “The development of a Sound Wheel for Reproduced Sound”, in *the 138th Conv. Audio Eng. Soc.* , (May, 7-10, 2015),Warsaw, Poland, (Audio Engineering Society, New York), Prepr. No 9310.

[53] P. Legendre and L. Legendre, “Numerical ecology”, Numer. Ecol. Second English Ed. **20**, pp. 378-85 (1998).

[54] D. Davis and C. Davis, “The LEDE-Concept for the Control of Acoustic and Psychoacoustic Parameters in Recording Control Rooms”, J. Audio Eng. Soc., (9),pp. 585-595 (1980).

[55] S. E. Olive and F. E. Toole, “The Detection of Reflections in Typical Rooms”, J. Audio Eng. Soc. **37**, pp. 539–553 (1989).

[56] T. Hidaka, L. L. Beranek, and T. Okano, “Some considerations of interaural cross correlation and lateral fraction as measures of spaciousness in concert halls. Chapter 32”, in *Music Concert Hall Acoust.*, edited by Y. Ando and D. Noson (Academic Press, London, UK) (1997).

[57] S. Hameed, J. Pakarinen, K. Valde, and V. Pulkki, “Psychoacoustic Cues in Room Size Perception”, in *Proc. 116th Conv. Aud. Eng. Soc.*, May, 8-11, 2005, New York, USA, (Audio Engineering Society, New York), Prepr. No 6084.

[58] B. Soerensen, “US Patent: 7702123 B2 - Loudspeaker assembly”, (2010).

[59] P. Hegarty, M. Møller, M. Olsen, M. Lydolf, and J. A. Pedersen, “Room gain in a car”, in *48th Int. Conf. Audio Eng. Soc.*, Munich, Germany, September, 21-23, 2012, (Audio Engineering Society, New York).

[60] K. Beresford, “Perceptual Effects of Spectral Magnitude Distortions in a Multi-Channel Automotive Audio Environment”, Ph.D. thesis, IoSR: University of Surrey (2010).

[61] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, “Elicitation of attributes for the evaluation of audio-on-audio interference”, J. Acoust. Soc. Am. **136**, pp. 2630–2641 (2014).

[62] S. Le Dien and J. Pagès, “Hierarchical multiple factor analysis: Application to the comparison of sensory profiles”, Food Qual. Prefer. **14**, pp. 397–403 (2003).

Paper D

On the Perception and Preference of Reproduced Sound in Ordinary Listening Rooms

Neofytos Kaplanis, Søren Bech, Tapio Lokki, Toon van Waterschoot, and Søren Holdt Jensen.

The paper has been submitted to the
Journal of Acoustical Society of America. Currently under peer-review.

The layout has been revised.

Abstract

An experiment was conducted to identify the perceptual effects of acoustical properties of ordinary listening rooms in a sound reproduction scenario. Nine sound fields, originating from four standardized residential listening rooms were captured and spatially reproduced over a 3D loudspeaker array. A panel of ten expert assessors identified and quantified the perceived differences, using own perceptual attributes. Following a multivariate analysis, two principal components summarize the sound fields of this investigation, relating to the decay times of the enclosure and the low frequency energy content. Four perceptual constructs seem to characterize the sensory properties of these dimensions, relating to Reverberance, Width & Envelopment, Proximity, and Bass. Overall, the results signify the importance of reverberation in ordinary listening rooms on the perceived sensory experience and in consequence, the assessors' preferences.

1 Introduction

In a typical real life scenario, where a listener attends to a sound source in an enclosed space, the perceived sound is a mixture of the emitted sound and a multitude of delayed and modified copies of it. This phenomenon occurs as the propagating wave interacts with the enclosure's boundaries, producing a distinctive sound field; what is commonly referred to as *reverberation*. The effects of reverberation on the perceived aural experience are well investigated in performance spaces [1–3] and form anticipated characteristics for their designers [4].

The direct investigation of these properties in geometrically smaller spaces, such as residential ordinary listening rooms, is limited. The physical dimensions of these environments are restricted compared to performance halls, exhibiting lower values of *Reverberation Time* (RT), denser early reflection patterns, as well as distinct and strong first reflections. In addition, their restricted size introduces a dominant modal behavior at lower frequencies, creating spectral irregularities within the enclosure. The source is typically a sound reproduction system, encompassing its own properties, which interact heavily with the room [5, 6]. This interaction influences both the timbral [7, 8] and spatial [9] characteristics of the perceived listening experience , as well as listeners' preference [10]. It is therefore unclear to what extent findings in other domains, e.g. in performance spaces, could be used to characterize the perceived aural experience in the smaller, ordinary listening rooms [11].

In this study, we seek to investigate the effects of commonly encountered acoustical variations of such rooms on the perceived sound field, in a typical sound reproduction scenario. Understanding the properties of human perception in ordinary rooms and its relation to physical characteristics of the field, would enable a more faithful reproduction of the recorded signals, aided by perceptually-relevant room compensation algorithms. This may increase the capacity of reproduction systems to perceptually recreate divergent kinds of sound fields and soundscapes in domestic environments, a highly anticipated property for today's multichannel rendering schemes [12, 13].

A wealth of scientific investigations in sound reproduction [5] and spatial audio [12] attempted to decode the mechanisms that dominate the perception of sound in ordinary rooms. Driven by the precedence effect [14], research was steered towards the audibility thresholds [15] and perceptual relevance [5, 6, 9] of distinct, early reflection patterns. Still, these investigations focused on the interaction between the loudspeakers' characteristics and the room, rather than the perceptual influence of the acoustical properties of the reproduction room on the perceived sound.

The aims of the current study are: (1) investigate the extent to which common acoustical properties (e.g. decay times) of ordinary rooms affect the perceived sensory experience, (2) identify the major perceptual attributes underlying these properties and the relationships between them, and (3) examine possible influences of the physical and sensory characteristics of these fields on assessors' preferences.

In Section I.A, the rationale behind the experimental design is stated. Section II presents the experimental methods, including the experimental conditions, the apparatus, and the evaluation procedures followed. The statistical analysis of the experimental results is then described in Sec. III. The findings are further discussed in Sec. IV, including general remarks and limitations of the study. Finally, in Sec. V the study is summarized and concluding observations are given.

1.1 Background & Motivation

Conducting perceptual evaluation of room acoustics is not a trivial matter [11, 16]. Challenges regarding the acquisition, presentation, and evaluation of acoustical fields [17], as well as potential perceptual biases of human assessors, restrict the capacity of repeatable and scientifically-valid experimental frameworks [18]. In consequence, the methodology followed is typically driven by the contextual factors, limitations, and proposed applications, resulting in domain-specific approaches.

In practice, one could seek the perceived differences between room simulations, allowing high parameter control over physical alterations of the fields. The generality of such investigations to real spaces, however, is challenging [19]. *In-situ* evaluation of reproduced sound in ordinary listening rooms has also been conducted [20]. It is however impractical and requires long time intervals between comparisons. It was recently indicated that even a few seconds between comparisons altered assessors' judgments [21, 22]. These findings relate to the limited auditory memory that humans exhibit [23] and the low level multi-modal processes associated with *room adaptation* [24], where the assessors' psychological state and perceptual sensitivity is modified [11]. Thus, the extent to which repeatable and robust results could be observed in such evaluation protocols is limited.

In room acoustics assessment this limitation could be avoided by employing a presentation scheme where the sound fields under investigation are evaluated simultaneously, in blind experimental designs [11, 17]. In this experiment, the perceptual evaluation was conducted in a laboratory, where human assessors were presented with auralized sound fields [25] based on *in-situ* measurements of real enclosures. The measurements included acoustical fields captured in several listening rooms and settings, and they were spatially reproduced over a spherical loudspeaker array in an anechoic chamber. This allowed controlled, blind, and instant comparisons between diverse acoustical properties. Similar approaches have been successfully applied in audio evaluation of the perceived timbral and spatial characteristics of acoustical properties of concert halls [3, 16, 26], critical listening environments [27], and car cabins [17, 28].

When evaluating real life soundscapes, that are well familiar to human assessors, one should control possible biases relating to past experiences [24, 29] and internal references, e.g., memorized schema of a typical stereo reproduction [12, 30]. Such biases may influence assessor's judgments relating to own internal references and affective biases. In order to overcome this, it is recommended [18] to conduct experiments using uni-dimensional and quantifiable features of the aural experience, referred to as *perceptual attributes*. Here, individually elicited perceptual attributes were used in the experimental design, based on a rapid *Sensory Analysis* (SA) [31, 32] framework. This family of protocols allows the sensory characterization and quantification of complex and multidimensional stimuli, avoiding global and hedonic biases. These techniques have been instrumental in audio evaluation studies, such as spatial audio reproduction through loudspeakers [33, 34], concert halls [3, 26], and automotive audio [28, 35].

2 Methods

The experimental design followed the *Flash Profile* (FP) [36, 37] protocol, adapted for evaluation of audio material [17]. Previous studies [17, 28] discussed the rationale and the practical implementation of the experimental methodology in detail. In this section the applied methodology is briefly presented.

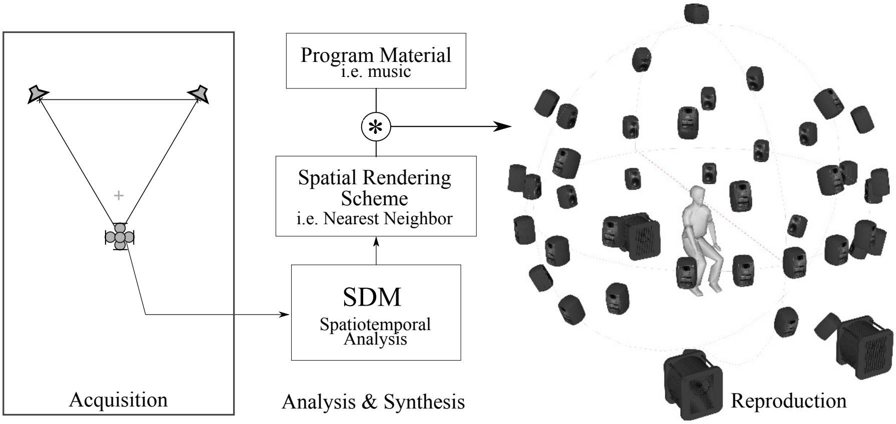


Fig. D.1: A schematic describing the processes followed during acquisition, analysis, synthesis, and presentation of the sound fields.

2.1 Acquisition of Room Impulse Responses

In-situ measurements were conducted to obtain spatial *Room Impulse Responses* (RIR). The excitation sources, two full-range loudspeakers (Genelec 1031A with 7041A subwoofer), were placed in a 2-channel stereophonic configuration [38, 39]. The sources were level-matched individually and the total system was calibrated at the listening position, at 82dB (C-weighted) sound pressure level, using 10 s of pink-noise.

The acoustical field was captured by a 3D vector intensity probe (G.R.A.S, 50VI-1), comprising of two coincidental microphones on each axis, separated by 25 mm. The RIR of each loudspeaker was measured separately, using a 5 s logarithmic sine sweep [40] at 192 kHz sampling rate.

The captured RIR were analyzed using *Spatial Decomposition Method* (SDM) [25]. SDM provides a spatiotemporal analysis, where the instantaneous direction of each discrete sample is calculated. The spatial information is then combined with the captured pressure, to spatially synthesize the acquired sound field. That is, the acoustical energy is distributed to a set of reproduction loudspeakers using a parametric reproduction scheme; here the nearest neighbor technique was followed [41]. The sound field acquisition, analysis, and reproduction schemes are graphically summarized in Fig. D.1.

2.2 Room Acoustic Parameters

During the acquisition of the RIR, described above, an additional set of measurements was performed for the calculation of standard acoustical parameters [42, 43]. To achieve precision RT measurements [43], the acoustical energy was emitted by an *OmniSource* 4295 (Brüel & Kjær) and a subwoofer YST-W300 (Yamaha), at three positions, and captured by four arbitrarily placed microphones (1/2” 40AZ, G.R.A.S. Denmark). The acquisition of the RIR was achieved by the sine-sweep method [40], sampled at 48 kHz. The estimation of RT_{30} was based on Schroeder’s backwards integration method [43] on theRIR, averaged across ten repeats for each source-receiver combination [44].

2.3 Experimental Conditions

The IEC:60268-13 [38] recommendation describes a set of properties and settings, that represent the acoustical and physical characteristics of residential listening environments. It includes general constraints in the physical dimensions of the enclosure and relative ratios between boundaries, as well as a range of possible RT across frequency. A large survey of residential spaces also indicated similar findings [45].

Four listening rooms were used in the experiment, representing a set of residential, ordinary listening environments. All four rooms, labeled as Room A–D, comply with the size and structural requirements of IEC:60268-13 [38]. The structural and acoustical characteristics details are given in Table D.1.

Rooms A, B, and D, comprise of wooden floors and ceiling, and acoustically treated boundaries, including ordinary furniture, and fully comply with IEC:60268-13 [38] specifications. Room C, is a critical listening space [46], built to host multichannel reproduction layouts [39]. It is therefore characterized by low RT_{30}, larger volumetric size than a typical room, and does not include furnishing.

In order to capture a wide range of possible acoustical fields, the interior of Room D was physically modified. The modifications aimed to vary the RT of the field, as uniformly as possible across frequency. The modular structure of Room D allowed successive modifications in all boundaries. Modifications in the lateral plane were completed in symmetry. The reproduction system was fixed and the direct paths between the source, ceiling, floor, side walls, and the receiver were not acoustically modified. This ensured constant contribution of the first reflection points in these measurements, as shown in Table D.1.

The reproduction system was positioned at a common relative point in all rooms. This was followed to avoid extreme alterations at low frequencies, due the rooms’ dissimilar modal regions [5]. Figure D.2 depicts the overlaid diagrams of the setup during the acquisition of RIR in all rooms. The measured sound pressure at the listening position is shown in Fig. D.3a.

2.4 Stimuli Selection

Nine acoustical conditions were selected for the listening experiment. That is, the unmodified sound fields of Room A, B and C, and a set of six acousti-

Table D.1: Physical properties of the acoustical conditions used in the experiment. Acoustical parameters were calculated based on ISO:3382 [42, 43]. The first reflection points were calculated using spatiotemporal analysis [47].

Condition	V[m^3]	RT30$_{0.5-1k}$ [s]	EDT$_{0.5-1k}$[s]	DRR	TS[s]	C50[dB]	Side [dB]	Floor [dB]	Ceiling [dB]	Compliance
Room A	80	0.32	0.34	-7.49	0.28	0.55	-9	-6	-5	IEC:60268-13 [38]
Room B	62.4	0.36	0.37	-8.90	0.24	1.22	-12	-8	-6	IEC:60268-13 [38]
Room C	172	0.17	0.11	3.64	0.05	21.90	-25	-7	-14	ITU-R:1116-2 [39] & [38](size only)
Room D-1	90	0.30	0.29	-4.33	0.17	5.85	-9	-8	-10	IEC:60268-13 [38]
Room D-2	90	0.43	0.39	-6.22	0.24	3.09	-9	-8	-10	IEC:60268-13 [38]
Room D-3	90	0.50	0.46	-8.92	0.33	-0.37	-9	-8	-10	IEC:60268-13 [38]
Room D-4	90	0.65	0.56	-9.79	0.38	-1.49	-9	-8	-10	RT$_{500Hz}$ > IEC:60268-13 [38]
Room D-5	90	0.67	0.61	-12.25	0.45	-2.66	-9	-8	-10	RT$_{250-4k\ Hz}$ > IEC:60268-13 [38]
Room D-6	90	0.83	0.72	-13.05	0.53	-3.99	-9	-7	-9	RT$_{63-8k\ Hz}$ > IEC:60268-13 [38]

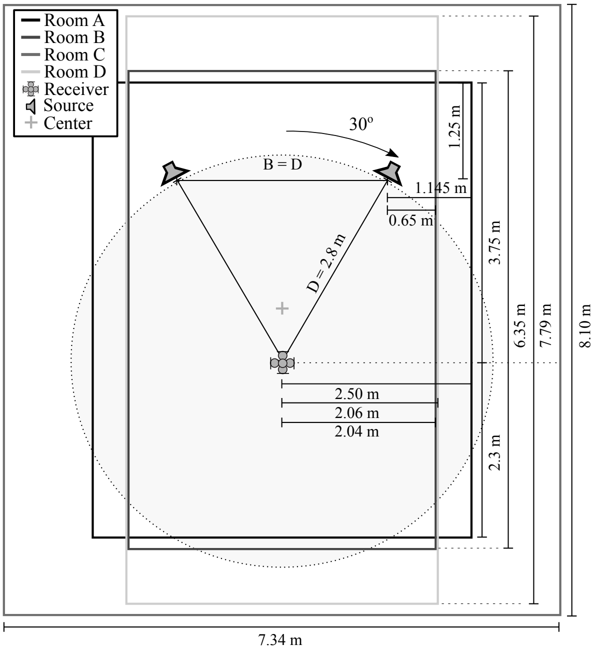


Fig. D.2: Experimental setup used in the measurements, including the dimensions of the four rooms used in this study, and the relative position of the reproduction system. The overlayed diagrams are in scale.

cal conditions originating from Room D. The selection was based on a pilot experiment that assessed the ability of three expert assessors to discriminate differences between conditions. The final acoustical conditions provided a range of possible RT of an ordinary room as described in IEC:60268-13 [38], including extreme cases. They are labeled as Room D-1 to D-6, ranging from the lowest measured RT_{30} to the highest, respectively. The nine acoustical conditions used in the experiment formed the levels of the first *Independent Variable* (IV), summarized in Table D.1. Their calculated RT_{30} are graphically shown in Fig. D.4.

To reproduce the acoustical conditions in the laboratory, the analyzed spatial RIR were convolved with the three program materials, loudness–matched before convolution at -15 dB_{LUFS}. The program materials included: (1) *Shola Ama – You might need somebody, 0:09–0:25*, (2) *Anechoic Kongas African Rhythms – Music For Archimedes [48], 0:24 – 0:33*, (3) *Female Speech English – EBU SQAM [49], 0:00–0:15.* These three excerpts were selected based on the ecological validity given for the experimental scenario, their dissimilar spectral, dynamic, and spatial properties, shown in Fig. D.3b, and their ability to excite the differences between the acoustical conditions selected. These excerpts formed the three levels of the second IV, the Program, and are further referred

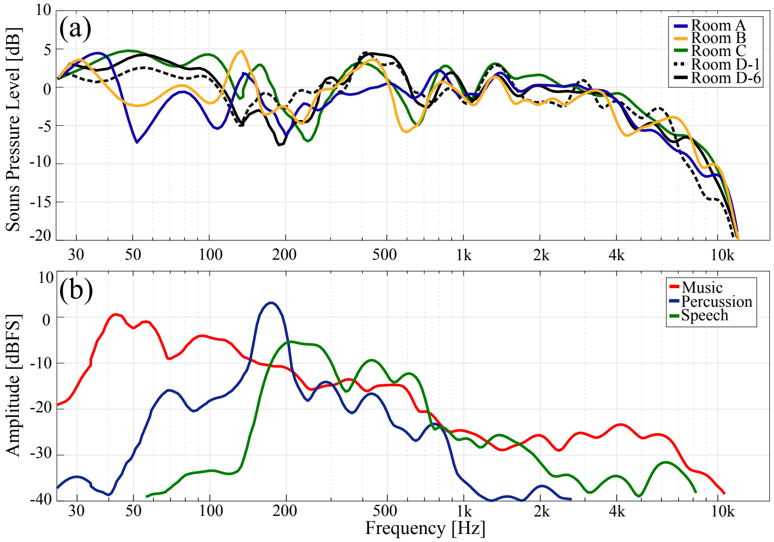


Fig. D.3: The magnitude response across frequency for: (a) the pressure at the listening position, and (b) the program types used in the experiment. Levels were normalized and 1/3 octave averaging was applied for visualization purposes.

to as music, percussion, and speech, respectively.

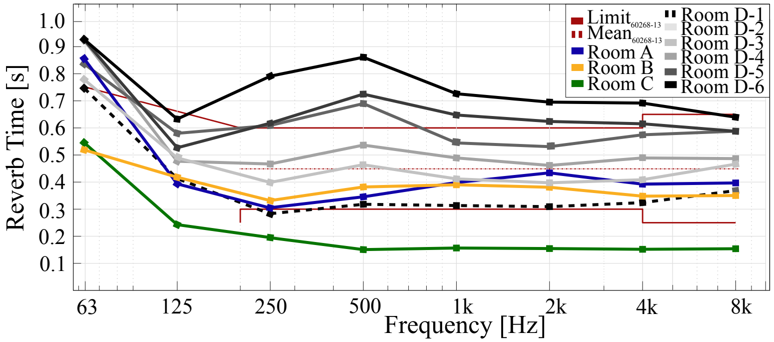


Fig. D.4: The Reverberation Time (RT_{30}) of the sound fields used as the experimental conditions in the study, calculated in octave bands. The recommended RT tolerances of IEC60268-13 [38] are depicted.

2.5 Experimental Apparatus

Forty full-range loudspeakers (Genelec 8020C) and three subwoofers (Genelec 7050B) were placed in an anechoic chamber^a, in a spherical orientation of 1.55 m radius. The physical placement of the reproduction loudspeakers in the laboratory was based on a spatiotemporal analysis [47] of the RIR captured in four standard listening rooms, previously described in Sec. II.C. The reproduction array is depicted in Fig. D.1. The tolerance between the individual loudspeakers was ±1.5dB at 25–16k Hz. All forty-three loudspeakers were temporally matched at the listening position, to account for any inherent inconsistencies of their positions.

To avoid possible visual biases, the assessors were guided in the anechoic chamber by the experimenter under dark conditions. The experimental setup included an acoustic curtain that separated physically and visually the assessors and the experimental apparatus. During the experiment the reproduction level was 75 ± 0.5 $\mathrm{dB}_{\mathrm{LAeq(15s)}}$ at the listening position, across all stimuli.

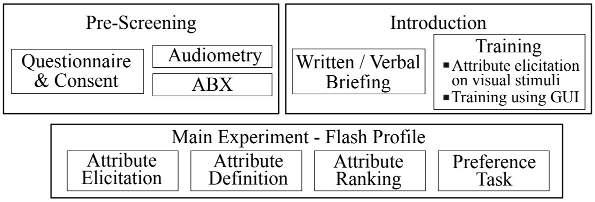


Fig. D.5: Experimental procedure.

2.6 Experimental Procedure

The experiment was completed in a single session and comprised of three phases: pre-screening, introduction, and the main experiment, the FP protocol. The experimental procedure is summarized in Fig. D.5.

First, the assessors completed a questionnaire, where background information was collected and consent was given. During pre-screening, the assessors' hearing sensitivity was evaluated [50]. Their ability to discriminate differences between alterations of room acoustical parameters was then assessed, over a discrimination experiment (ABX). The assessors were presented with three stimuli on each trial, labeled as 'A', 'B', and 'X'. Their task was to identify which of two choices was the most similar to stimulus 'X'. A set of sixteen acoustical conditions and two program types were evaluated. The stimuli included in this short screening experiment were created using identical procedures as described above, and presented on headphones, using non-individualized binaural synthesis. In order to avoid familiarization and adaptation effects, the stimuli-set made use of similar acoustical fields as the final experiment, i.e., the relative RT differences between the conditions was similar to the experimental conditions (Sec. II.C), but not the actual sound fields or program materials. Only assessors that achieved more than 75 % correct responses continued to the next phase.

The assessors were then introduced to the final experimental procedure, the principles behind FP and the experimental methodology, in verbal and written form. An attribute elicitation procedure was then followed, for a selection of visual stimuli. A set of dissimilar objects was presented on screen and the assessors were asked to characterize their differences using uni-dimensional, scalar, and non-hedonic descriptors. Finally the assessors familiarized with the user interfaces on a tablet.

This main listening experiment included four parts, the attribute elicitation, the attribute definition, the ranking, and the preference task. These tasks were conducted in the experimental apparatus described in Sec. II.E, under dark conditions. The assessors were asked to provide as many verbal descriptors as necessary to characterize the perceived differences between the presented sound stimuli. They were allowed to label the extreme intensities of that attribute, e.g., for the attribute 'Bassiness' one could set high intensity to 'too much' and low intensity to 'just audible'. During the elicitation procedure, all stimuli were available on screen, including the nine acoustical conditions and the three program materials.

Following the attribute elicitation, the assessor provided definitions for the given attributes and ordered them based on the perceived audibility and perceptual importance. Then, the assessors performed the ranking phase of the FP protocol. A block design was followed, where each block evaluated one attribute, in three trials, one for each program type. At each trial, the assessors were exposed to ten stimuli, the nine acoustical conditions, and a hidden repeat. The hidden repeat was included to assess the ability of the assessor to correctly discriminate and quantify the stimuli on the given trial. The scales were presented as a continuous vertical slider, 15 cm long [31]. Assessors were able to loop any time segment of the presented stimulus, as in standard evaluation procedures [51]. The presentation order was counterbalanced by randomization for all experimental variables and assessors.

After successful completion of the FP procedure, assessors were asked to indicate their personal preference to the presented stimuli. Prior to this phase, assessors were not informed about the context of the presented sound fields, nor the preference task, to allow faithful and unbiased judgments between the stimuli [18]. For the preference task however, it deemed necessary to inform the assessors about the experimental details, so that an ecologically valid scenario was realized and the appropriate hedonic response was evoked. Literature pertaining the evaluation of real scenarios in the laboratory [34, 52] suggests to envisage assessors to a given situation. In the current study, the assessors were given the instruction of: "Imagine that you are in a living room listening to a 2-ch stereophonic reproduction over loudspeakers. Please rank the presented stimuli in a way that expresses your preference from 'highly dislike' to 'highly like'." The attribute 'Preference' and its anchors were presented on screen, and assessors performed three trials, one for each program type.

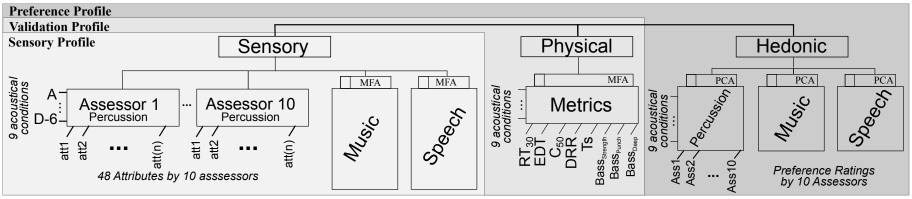


Fig. D.6: Data structure used in the statistical analysis. Each hierarchical node, denoted as Sensory, Physical, and Hedonic, describe the observations of a given dataset, used as the main hierarchical nodes in *Hierachical Multiple Factor Analysis* (HMFA). The horizontal lines identify the hierarchical nodes for a given hierarchy tree. Sensory data included the individually elicited attributes, used for calculating the sensory profile in Sec. III.A. Physical data included Reverb Time (RT_{30}) and Early Decay Time (EDT) for each of the 31.5-8k Hz octave bands, Clarity Index (C50), Direct-to-Reverb-Ratio (DRR), and three metrics for the perceived bass, namely the Bass -Strength, -Punch, and DeepBass. The sensory and physical data were combined for calculating the validation profile, in Sec. III.D. Hedonic data include the preference ratings of each assessor, grouped by program type. All three datasets were combined to provide a global preference profile, presented in Sec. III.E.

During all the experimental phases assessors followed self-paced and self-controlled procedures, using a custom MAX/MSP GUI on a tablet [17]. No time limits were set to complete the tasks but short breaks were regularly completed to avoid possible listening fatigue. The whole experimental procedure, including breaks, was completed in 55 min–125 min (Mean = 84 min).

2.7 Assessors

Ten assessors participated in the experiment as volunteers. The assessors are considered as experienced listeners, with an average experience of 11.4 years (*s.d.* ± 5.5) in acoustics research and development, spatial audio reproduction, and critical listening as part of their profession. All assessors reported proficiency in standard audio evaluation procedures. Seven assessors were familiar with sensory analysis protocols for evaluation of audio material and six had performed attribute elicitation procedures before. They were all male, aged between 27-47 years old (Mean = 34.1, *s.d.* ± 6.2). Their hearing sensitivity was confirmed to be above 20 dBHL between 125–8k Hz by standard audiometric evaluation [50].

3 Results

In the current experimental design nine acoustical conditions were evaluated in terms of their sensory and hedonic characteristics for three program types. All ten assessors correctly identified the hidden repeats, in both preference and sensory tasks and no data points were eliminated for the statistical analysis. The experimental *Dependent Variables* (DV) included 48 individually elicited attributes, describing the perceptual differences between the stimuli. These observations are further referred to as *sensory data*. The assessor's hedonic responses were also collected in a free preference task, and they are further referred to as *hedonic data*. The physical properties of the sound fields were also calculated, and will be referred to as *physical data*. The datasets are summarized in Fig. D.6.

In classical SA, i.e., of food products [31], the statistical analysis of such data is achieved by utilizing multivariate techniques such as *Generalized Procrustes Analysis* (GPA) [53] and *Multiple Factor Analysis* (MFA) [54]. When evaluating audio material, however, the stimulus-under-test is always the product of an acoustical condition, e.g., spectral modification, and an excitation program, e.g., speech [17, 18]. This property of audio evaluation is limiting, as possible interactions between the two factors, i.e. the acoustical conditions and the excitation programs, may not be easily identified by common univariate techniques. To overcome these challenges, an extension of MFA was followed in this study, namely, the *Hierarchical Multiple Factor Analysis* (HMFA) [55].

HMFA applies a multivariate analysis to a group of variables that are hierarchically inter-related. In a given dataset, HMFA performs a factorial analysis at each hierarchical node of a defined hierarchy, while keeping the role of each node equal to the global factorial space. This enables a balanced statistical analysis of several multi-table datasets, aiming to identify the most prominent components between them and their underlying relationships. In an audio evaluation protocol, it allows the analysis of the perceptual dimensions for all assessors in a common factorial space, while the within-variance, i.e., the within-inertia, of each excitation program is individually computed and preserved.

3.1 Ordination of Sensory Data using HMFA

Initially, an HMFA was conducted [56] on the sensory data. The three hierarchical nodes of the sensory data included the observations of the nine acoustical conditions, for each program type separately, i.e., speech, music, percussion, as shown in Fig. D.6. This analysis provided a global factorial space of the assessors' perceived differences and similarities for the presented acoustical conditions. Yet, the observations were analyzed for each node separately, i.e. each level of the program, and a common solution was calculated. This analysis allows the evaluation of the main effects of the design's IVs. That is, the acoustical Condition and Program.

Similar to its predecessor, the MFA, HMFA provides a graphical ordination of the stimuli in a common factorial space, known as *factor map*. Figure D.7 depicts the factor map of the sensory data, including the partial contributions of the hierarchical nodes, i.e., the program levels. The partial points for each condition denote the direction of the variance, explained by the three data sets, i.e. the observations when listening to music, speech, and percussion.

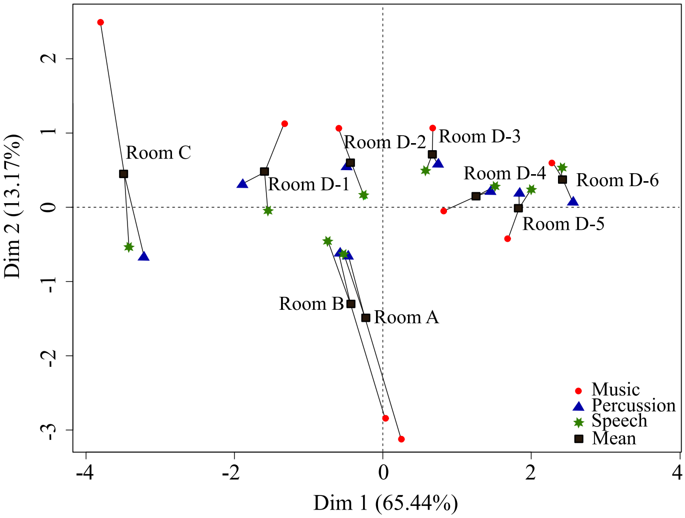


Fig. D.7: Factor map showing the sensory profile of the experimental conditions. The profile is a result of three separate analyses, one for each program type. The mean is then calculated based on equally weighted contributions via the HMFA analysis, and the individual components are visualized in a common factorial space as partial points.

The barycenter of these points, gives the mean ordination of the individual acoustical condition, across all programs.

The first two dimensions represent adequately the data, accounting for almost 80% of the total variance. All acoustical conditions are adequately separated, and no overlaps can be seen. On the first dimension, Room C and Room D-6 hold the extreme points and the remaining conditions are ordered in between these. It is noted, that acoustical conditions with low RT_{30} values are positioned negatively on this dimension, and the more reverberant conditions, on the positive side. One could hypothesize that this dimension may relate to the decay time of the acoustical conditions, as the position of the conditions matches their calculated RT and *Early Decay Time* (EDT), i.e., from the lowest value, Room C, to the highest Room D-6.

The second dimension is mainly driven by Rooms A and B, while the remaining acoustical conditions share similar coordinates on this dimension. However, in some conditions, e.g., Room A, Room B, and Room C, a strong opposition between the different program types is apparent, especially on the second dimension. To further analyze this interaction, the axial inertia for each of the hierarchical nodes for the first five principal components of the HMFA solution was calculated, shown in Table D.2. These results indicate that although the solution of HMFA was balanced across program types for most dimensions, the second dimension is highly driven by the assessors' ratings relating to music, compared to ratings when assessors listen to speech, and percussion. This suggests that music ratings were more sensitive to the perceptual property underlying the second dimension; compared to when percussion and speech were used as the excitation signal. This could be related to the properties of the program materials when combined with the acoustical conditions used.

Table D.2: Decomposition of inertia of the HMFA analysis, in eigenvalues, split by each program type.

	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5
Percussion	0.99	0.09	0.13	0.05	0.04
Music	0.95	0.43	0.09	0.09	0.07
Speech	0.99	0.07	0.05	0.06	0.05
Total Inertia	2.93	0.59	0.28	0.20	0.17
Total Inertia %	65.44	13.17	6.16	4.48	3.70

3.2 Clustering of Individually Elicited Attributes

In the above analysis the sensory profile of the presented acoustical conditions has been summarized. This factorial solution depicts the ordination of the experimental stimuli based on the perceived differences and similarities of all assessors, i.e., the consensus sensory profile. It will be more informative, however, if the perceptual properties underlying these factorial relationships are included in the same analysis [3]. This is a trivial matter when assessors use a common list of attributes [57], known as a consensus vocabulary, as each attribute is a common variable across all assessors [31].

In individual vocabulary methods, such as FP, each assessor elicits and uses own attributes. Thus, the semantic meaning of the collected attributes, their anchors, and intensity levels are unique to each assessor. Naturally, this limits the extent in which one could summarize the perceptual dimensions of all assessors by assuming semantic equivalence.

In order to identify the common *perceptual constructs* underlying the individually elicited attributes across all assessors, a *post-hoc* analysis is required. This is typically achieved by grouping the individually elicited attributes into common semantic categories, i.e., based on the given definitions [34]. It was however shown that the classification of the attributes into collective categories could be based on the true structure within the collected data [16, 17]. This may limit potential interpretation biases, especially in individually elicited vocabularies, where each assessor labeled, defined, and rated the perceived differences with no guidance.

Here, the grouping of the collected attributes was achieved using *Agglomerative Hierarchical Clustering* (AHC), similar to previous individual vocabulary studies [3, 28]. An AHC based on the Euclidean distances of the MFA coordinates of each attribute was calculated, in conjunction with Ward's criterion [58]. As AHC assigns equal weight to all attributes, independent of the contribution to the principal components, any individual attribute that did not correlate well to any of the first two principal components was excluded [28], i.e., $| r |_{Dim1/Dim2} < 0.65$ [b]. A total of 43 attributes were used in the AHC analysis, as the best representative drivers to the dimensions of interest.

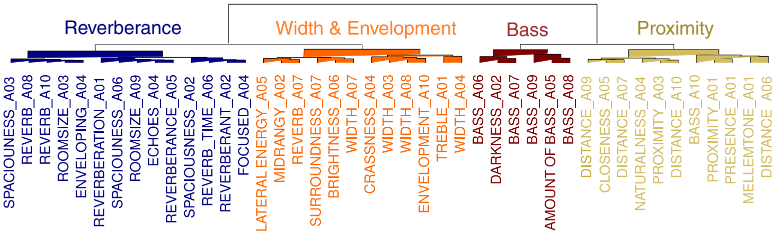


Fig. D.8: Clustering dendrogram of individually elicited attributes of each assessor, based on the MFA coordinates of each attribute. The assessor's number is also included for each attribute, denoted as 'A_{01-10}'.

Two main clusters were identified, comprising of four main groups of variables, as shown in Fig. D.8. The first cluster splits into two branches. The first branch highlights attributes relating to the perceived effects of *Reverberance*, e.g. $\text{Reverberation}_{A01,08}$, $\text{Room Size}_{A03,09}$. The second branch consists of attributes relating to the main components of spatial impression [59] of the sound field, i.e. the perceived *Width & Envelopment*. Lastly the second cluster clearly identifies attributes relating to the perceived *Bass* content, on its first branch. Its second branch includes attributes relating to the perceived distance, closeness and proximity. For consistency with previous studies [3, 60], this cluster will be refereed to as *Proximity* [11], indicating how close the auditory event is perceived[c].

It is noted that the semantic equivalence between the attributes within certain clusters may not be strongly inferred at this point. This is a common observation in free verbal elicitation experiments [57], where assessors label and quantify their perceived sensations without guidance or anchoring. As AHC seeks to classify interrelated attributes into a group, the attributes whose ratings were similar across the conditions, are classified as homologous. The clusters are then labeled based on the given attribute definitions [57], and the related literature [11]. A validation of these hypotheses follows, in Sec. III.D.

The four identified clusters form the perceptual constructs of this study, underlying the properties of the perceived differences across the assessors and will be further referred to as *Reverberance*, *Width & Envelopment*, *Bass*, and *Proximity*.

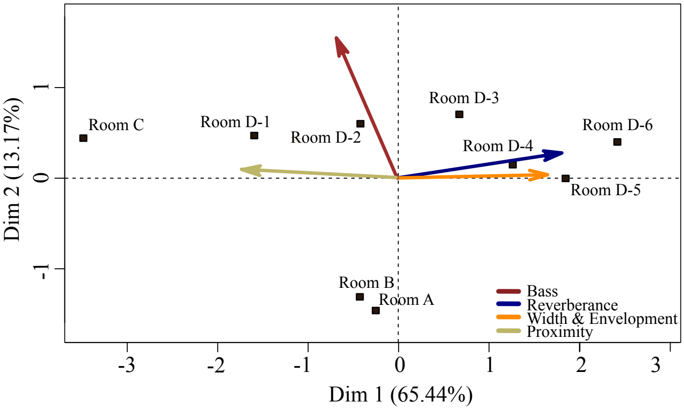


Fig. D.9: Biplot showing the sensory profile of the acoustical conditions. Vectors indicate the direction of inertia, given by the mean coordinate each cluster, i.e., the perceptual construct, and colored based on the identified clusters.

3.3 Sensory Profile of Acoustical Conditions

One could combine the two analyses, the HMFA ordination and the clustering results, in the form of a Biplot. This would allow a better understanding of the relationship between the perceived differences across the stimuli and the perceptual constructs underlying these relationships.

Fig. D.9 shows the ordination of the acoustical conditions, i.e., the mean coordinates calculated in Sec. III.A., and the projections of the perceptual constructs, identified in Sec. III.B, in the form of vectors. The direction of each vector indicates the direction of inertia within the conditions that is driven by the ratings of the clustered attributes, i.e. the perceptual construct. In effect, the projected vectors provide a perceptual explanation for the positioning of the acoustical conditions in this two-dimensional factorial space. The length of each vector indicates the quality of representation of the perceptual construct to the factorial space, i.e., its cumulative correlation to the principal components. In consequence, the projections with low correlation to the solution will be poorly represented due to their low contribution to the explained variance, thus, challenging to interpret.

The analysis summarized in Fig. D.9 suggests that the first dimension relates to the perceived *Reverberance* and *Width & Envelopment* while it opposes the perceived *Proximity* of the sound source. *Proximity* vector faces the opposite direction, which indicates its negative correlation to the dimension. This relationship is well known to room acoustics. That is, the perceived distance increases as a function of reverberation, commonly explained by the higher values of *Direct to Reverb Ratio* (DRR) [61].

Reverberance, that is the feeling of being inside a bigger, i.e. reverberant room^c, is correlated to perceived *Width & Envelopment* on the depicted dimensions. This suggests that the ordination of acoustical conditions on the first dimension is driven positively by these perceptual constructs. That is, a condition that has been described and quantified as more reverberant by the assessors, has also elicited a more wide and enveloping feeling, and *vice versa*. Still, the perceptual constructs were distinct to each other, and assessors were able to separately evaluate them.

Several studies [11] have shown that the perceived width and envelopment is related to the early energy originating from different directions [11]. One should note that in the current experiment, there was no intention to systematically vary the early reflection patterns in a specific manner, but rather compare sound fields between typical listening environments, and uniform spatial alterations of RT within such spaces, as shown in Table D.1. As the assessors' task was to identify and quantify the perceived differences between the stimuli, the perceived *Reverberance* and *Width & Envelopment* were determined and quantified as separate percepts. Yet, their ratings were proportional, as one would expect for the stimuli set presented. In consequence, the statistical analysis indicates that the attributes relating to these perceptual constructs were classified in two separate branches, yet the direction of inertia explained by these constructs in the common factorial solution is similar.

The second dimension is described by the perceptual construct of *Bass*. All variants of Room D share similar positions on this dimension, indicating perceptual equivalence to the perceived *Bass* content. Room A and Room B are perceived as less bassy, compared to the remaining acoustical conditions. One should expect such results, as low frequencies in small enclosures are mainly driven by the modal behavior of the enclosure, in consequence, their physical dimensions. Based on the current evidence, one could infer that this dimension describes the perceived low frequency content, in terms of spectral dissimilarity between the different rooms. This is also evident when contrasting the magnitude responses between the rooms, in Fig. D.3a, as Room A and Room B indicate the least energy levels below 100 Hz, compared to the other rooms.

3.4 Validation of Sensory Profile

In the previous analysis, the sensory profile was constructed based on the assessors' sensory quantification and tangible explanations of the profile were identified and projected to a factorial solution. This analysis identified the sensory characteristics of the presented acoustical conditions, e.g., their perceived differences and similarities, and constructed a factorial plane describing these relationships based on solely perceptual data, as given by the assessors.

Using HMFA it is possible to assess the *construct validity* of the applied experimental methodology. That is, the extent to which the identified perceptual constructs and sensory profiling relate to the physical characteristics of these fields.

To evaluate this, the sensory data, i.e., the perceptual responses given by human assessors, and the physical data, i.e., the acoustical parameters of the sound fields, formed the hierarchical nodes of the HMFA. This analysis compared the profile of the stimuli based on the sensory characteristics, as given by the assessors' responses, and the profile based on the physical properties of the acoustical conditions, for example the RT_{30} of an acoustical condition.

Fig. D.10 shows the ordination of the acoustical conditions based on this analysis. The explained variance of the analysis is 80 %, suggesting a well represented dataset. Moreover, the close positioning of the partial points indicate a good agreement between the physical profile and the sensory profile as they hold similar coordinates in both dimensions. The partial points on Room A, Room B, and Room C, seem to differ in the second dimension, denoting a slight disagreement of the two datasets for these conditions.

To further investigate this, the physical metrics and the perceptual constructs were projected into the factorial space in a Biplot, shown in Fig. D.11. The acoustical conditions are positioned in the factorial plane based on the mean coordinates of the sensory and physical data. The projected vectors indicate the directions of variance explained by the variables.

On the first dimension, the variance explained by the measured RT_{30} and the EDT of the acoustical conditions indicate an excellent relation to the perceptual construct of *Reverberance*. The metrics relating to the temporal distribution of energy in the rooms, i.e. the *Clarity index (50ms)* (C50) and DRR, indicate strong correlation to the perceived *Proximity*, opposing the perceived *Width & Envelopment*. *Center Time* (TS) [43] correlates well to the first dimension, indicating its inversely proportional relationship to perceived *Proximity* of the source, as well as it's direct relation to the physical measure of RT_{30}.

Based on the sensory profile, identified in Sec. III.A, it was hypothesized that the second dimension may link to low frequency content of the conditions, as suggested by the perceived construct of *Bass*. To objectively verify this hypothesis, a physical metric for the perceived bass content could be used. Recently, Volk et al. proposed a series of perceptually-based metrics, the Bass Punch [62], Bass Strength [63] and Deep Bass [64], aiming to assess the perceptual properties of broadband signals at low frequencies. These metrics have been included in the analysis, shown in Fig. D.11, suggesting that the second dimension relates to spectral differences at low frequencies between acoustical conditions.

As described above, when performing HMFA, the inertia within the individual hierarchical nodes could be projected in the global factorial space. Here, the sensory data included three sub-hierarchies, based on each program type, as shown in Fig. D.6. The partial vectors of each perceptual construct and program are also projected in the factorial solution, shown in Fig. D.11. These

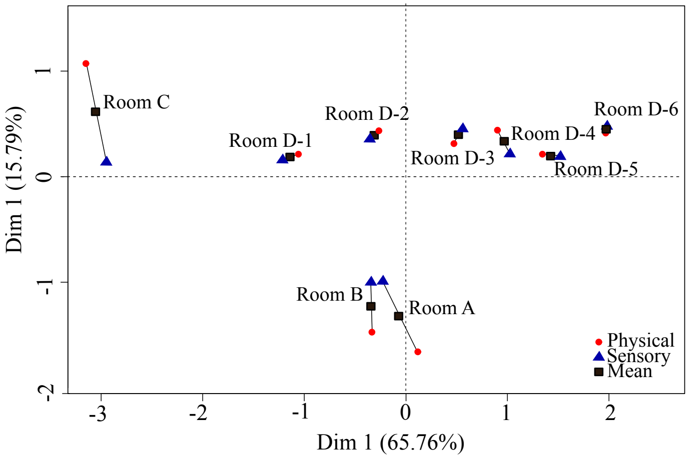


Fig. D.10: HMFA factor map, depicting the resulting profile of sensory and physical data. Partial points identify the ordination of the stimuli for the two hierarchical nodes of this analysis.

indicate the direction of inertia of each perceptual construct, when different program materials were used.

The previously identified interaction between the programs, seen in Sec. III.A, is also apparent here. The projection of the partial vector of *Bass* when assessors evaluated music content, suggests a good correlation to the second dimension and the factorial solution. In contrast the relatively short lengths of the partial vectors of *Bass* relating to the other two program levels, confirm a low quality of representation. This is an expected result, as the low frequency content of percussion and speech is limited, indicated in Fig. D.3b. It could be inferred that when these programs were used as an excitation signal, the audibility of these spectral differences between conditions was reduced, and in consequence the assessors' discrimination ability was affected.

In the previous section, a hierarchical clustering was employed to identify the common perceptual constructs. The clusters were then labeled based on the semantic definitions of assessors' own attributes and previous studies. These were projected into the factorial space, aiming to understand the perceptual relevance of the dimensions of the sensory profile. Here, the analysis suggests that the projections of the identified perceptual constructs share high similarities to the physical metrics that are known to excite such sensations [11].

In summary, this analysis suggests that the factorial space constructed from the assessors' quantification of the perceived differences relates to the physical characteristics of the sound fields used in this investigation, supporting the previous hypotheses, and the construct validity of the experimental design.

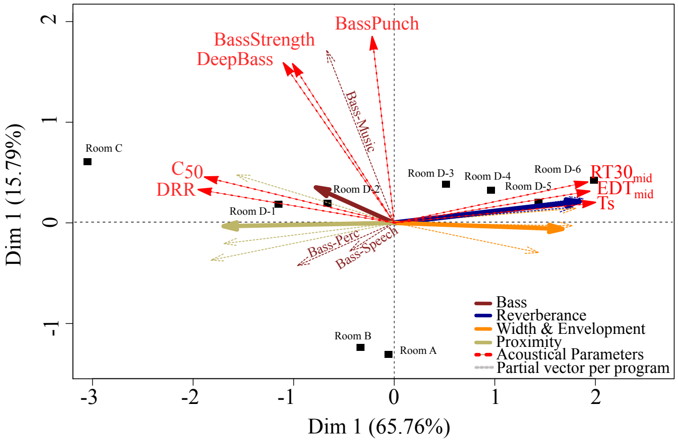


Fig. D.11: Biplot depicting the common profile for sensory and physical data, and the directions of inertia driven by variables in the sensory, and physical data. EDT$_{mid}$ and RT30$_{mid}$, refer to the average values between 500 Hz and 1k Hz as recommended [43] for perceptual relevance.

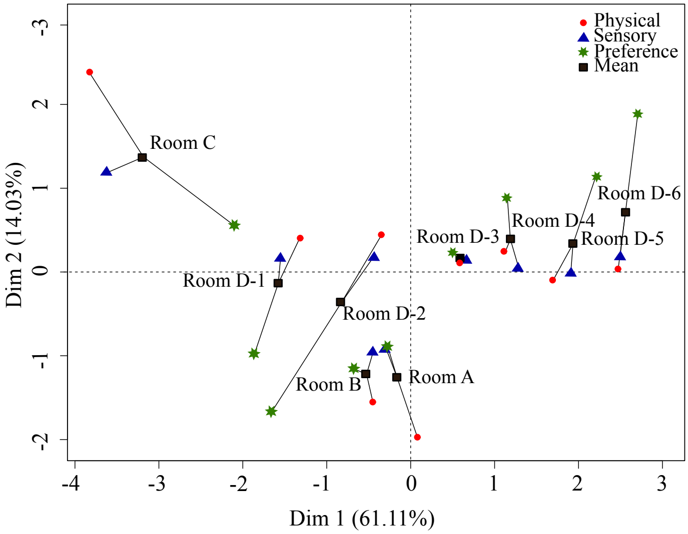


Fig. D.12: Preference factorial map for the consolidated profile of sensory, physical, and hedonic data.

3.5 Sensory, Hedonic, and Physical Data - Global Profile

In the final part of the experiment, the assessors' hedonic responses were also collected in a preference task. In order to understand the relations between the sensory, physical, and hedonic responses of assessors, *preference-mapping* techniques are typically followed [65]. Such analysis allows the direct identification of each acoustical conditions' sensory and physical characteristics, driving the assessors preferences.

For complex multi-table datasets, as the data in this study, preference mapping is achieved by computing the factorial solution of all the datasets simultaneously, and projecting the driving variables in that space [3]. Figure D.12 shows the common factorial space achieved by the three hierarchical nodes, the sensory, physical, and hedonic data, as described in Fig. D.6. The relative ordination of the acoustical conditions is similar to the previous analyses. The partial points shown in Fig. D.12 suggest a good agreement between the sensory and physical profiles. The hedonic profile includes larger differences mainly on the second dimension, yet, there is a good consensus to the first dimension.

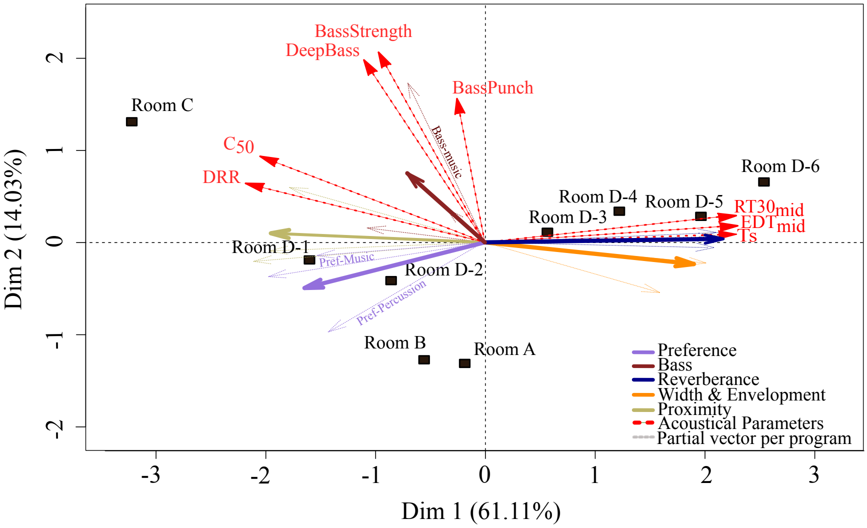


Fig. D.13: Biplot showing the common factorial space for sensory, physical, and hedonic profiles. Vectors indicate the direction of inertia for the variables from each profile.

This factorial solution could be explained by the Biplot shown in Fig. D.13. Here, the drivers explaining the variance in this factorial plane include both the sensory and the physical variables. The position of Room D-6 opposed the construct of *Preference*, suggesting that this condition was the least favorite across assessors, whereas Room C, D-1, and D-2 were the most preferred. The perceived sound in these sound fields exhibit high *Proximity* and they are identified as less reverberant, wide, and enveloping. Assessors' preference is explained well by lower decay times given by RT_{30} and EDT, higher levels of DRR and C50, and in consequence lower values of TS.

The perceived bass does not seem to explain the preference ratings adequately, yet, it indicates a contribution to the variance, due to its position on the first dimension. The depicted partial vectors, indicating the direction of inertia for a perceptual construct split by program type, have similar directions. That confirms that the mean coordinates of the perceptual constructs could explain adequately the factorial space, and no interactions between these partial programs can be seen in this analysis.

4 Discussion

The experiment investigated the perceptual effects of the acoustical properties of four ordinary listening rooms. In addition, a within-room analysis was included, where the listening room was physically altered to cover a range of possible decay times according to the IEC:60268-13 recommendation [38], including maximal anchors.

The primary findings suggest that the decay times of such fields is a dominant factor on the aural experience and it explains the majority of the perceived differences in this investigation. The RT of the sound field seems to influence the perceptual constructs relating to *Reverberance* and *Width & Envelopment*. In contrast, the perceived *Proximity*, that is the sense of a source being close or distant to the listener, is degraded at higher levels of RT.

In the current experimental design a controlled acquisition procedure was followed aiming to avoid strong modal behavior at low frequencies and focus on the spatiotemporal properties of the sound field. Still, the effects of low frequency differences between conditions, i.e. the four rooms, seem to be an important perceptual aspect; here labeled as *Bass*, which explained ∼13% of the variance of the sensory data.

The previously suggested two-dimensional character of the perceived differences in such environments [66] was also apparent here. The first identified dimension relates to reverberation and its cognate percepts, e.g, the perceived distance [67]. The second dimension relates to perceived spectral effects, as shown before [7, 68].

Four perceptual constructs seem to explain the factorial dimensions identified in this investigation, namely, *Reverberance*, *Width & Envelopment*, *Proximity*, and *Bass*. One could infer that this is a limited number of perceptual constructs and attributes, compared to concert hall [16] and spatial audio research [11]. It has however been noted [67], that a relatively few number of perceptual aspects are enough to describe small, ordinary rooms, where reverberation accounts for most. The current results indicate that assessors were able to identify and appropriately quantify the perceptual constructs independently^{d}. This suggests that although some constructs seem to correlate in the factorial solution, e.g., *Reverberance* and *Width & Envelopment*, they elicited different sensory aspects. It is therefore suggested that the four perceptual constructs identified in the study form major characteristics of perceived aural experience within the presented sound fields.

The preference profile, that combined the sensory, physical, and hedonic observations supports previous findings [27], that assessors systematically preferred the sound fields with lower RT. These fields evoked the sense of a less reverberant, and less wide and enveloping space. The sources were perceived as closer to the listener, exhibiting high levels of *Proximity*. It is remarkable to note that the current results suggested that a negative preference is apparent for acoustical conditions with RT higher than 0.4 s, the proposed mean value in the IEC recommendation [38]. Differences relating to the low frequency content within the presented conditions have also influenced assessors' preference, but at a lower degree than one could expect [69]; this may relate to the specific programs used in the study.

These results signify the effects of acoustical properties on the perceived sound and indicate the importance of strict RT limits in standard listening rooms, when loudspeakers [38] and impairments of audio signals [46] are evaluated. It was shown that slight alterations of the RT, e.g., ≈0.1 s between Room D-2 and Room D-3, evoked different hedonic and sensory percepts and intensities.

4.1 Limitations & Future work

Overall, the experimental paradigm seem to overcome multiple challenges relating to evaluation of room acoustics properties [11]. The assessors were able to identify and accurately quantify their sensations using a rapid sensory analysis method, the Flash Profile. Spatial aspects of the sound fields that one perceives in a real environment , i.e., distance from the source and apparent width, were evoked and appropriately quantified in a consistent way, compared to previous investigations that failed to reproduce such attributes [17, 67].

Nevertheless, the current approach has several limitations. FP requires all the stimuli to be available to the assessor at all times. In audio evaluation, this is a challenge. Typically several excitation signals (programs) are required to adequately assess the acoustical conditions of interest. A large variety of program material increases the ability to generalize the experimental results, yet the time required to complete the experiment increases significantly. It is therefore required that the experimenter balance this trade-off, by selecting the appropriate type and number of program materials.

Following the need of simultaneous presentation of all stimuli, another challenge is introduced. During attribute elicitation, it is possible to provide all program-conditions combinations in a single trial, as required. However, during the ranking phase, the experimental conditions must be evaluated for each program material separately [18], which requires a block design. Thus, although the individual attributes given by the assessors in the elicitation phase may have been evoked by certain program materials only, all attributes are used to assess all program-condition combinations during ranking. In consequence, the audibility of such differences might be reduced when other program materials are used in the ranking phase.

This may lead to two major implications. First the task becomes more difficult for the assessor and it requires more time and effort to complete. Second, it is likely to enhance statistical interactions between the acoustical conditions and the program materials. If one follows a simple multitable analysis such as GPA or MFA this interaction will result in an increase of the statistical error, hence decrease the explained variance, leading to misinterpretation of the experimental results. Here, HMFA was followed. The observations were analyzed for each program separately, and then they were combined in a common solution, preserving the contribution of each partial. This analysis depicted the perceptual construct of *Bass* to differ between programs and the statistical techniques followed accommodated this. It is therefore highly recommended to follow statistical methods that allow interactions between variables when analyzing similar datasets.

The study utilized a panel of expert assessors, trained in the domain of audio evaluation, as commonly recommended for FP studies. By conducting FP with expert listeners, the experimental time required is significantly reduced [17] and only few assessors are needed to complete the task, while no aural or vocabulary-based training is required [32]. Here, a clustering analysis categorized the individual attributes in common perceptual constructs with clear semantic meaning. This may relate to the robust and consistent quantification of the perceived sensations of experienced assessors, compared to a less trained panel [70].

However, this approach may have implications when evaluating the hedonic responses, such as preference and likeness. All assessors were familiar with standard listening rooms and critical listening environments due to their profession. In consequence, their internal inference and reference of a typical reproduction system may differ from an average listener. It is therefore noted, that the current findings relating to assessors' preferences may not reveal the judgments and sentiments of an average listener.

Further work is needed to expand the limited scope of the study. A spatial and temporal analysis of such sound fields could improve our understanding of the perceptual aspects of the fields, as identified here. This may include the investigation of the effects of early reflection patterns on the perceived sound from different directions, and the effects of irregular rooms and asymmetric reproduction setups. Moreover, the influence of the excitation signals seems to affect the perceived experience, and a detailed investigation should follow. Finally, future work will investigate sensory sensitivities and hedonic responses of non-trained assessors to identify the extent to which the the current findings can be generalized.

5 Conclusions

The paper reported an investigation on the perceptual effects of the acoustical properties in residential listening rooms. The acoustical conditions aimed to provide the assessors with scenarios they are likely to experience in real life. Nine sound fields, originating from four standard listening rooms, were evaluated by ten expert assessors. The sound fields were captured *in situ* and reproduced over a spherical loudspeaker array in an anechoic chamber. A blind experimental protocol was followed, where expert assessors identified own perceptual attributes and quantified their perceived sensations for the presented sound fields, with no prior information about the stimuli in a highly controlled experimental apparatus. Using multidimensional factorial analysis, a sensory profile for the acoustical conditions was calculated, depicting the perceived differences between the acoustical conditions. The main drivers of this profiling, i.e., the common perceptual constructs, were identified and further validated against physical properties of these sound fields. The assessors' hedonic responses were collected and analyzed indicating the assessors' preferences for a typical reproduction setup, in an ordinary listening room.

Overall, the current results suggest the dominance of RT on human perception in ordinary rooms, supplemented by spectral modifications at the low frequencies. As a result, the perceived *Reverberance*, *Width & Envelopment*, and *Proximity* seem to explain the majority of perceptual differences between the sound fields used in the experiment, in addition to alterations of the perceived *Bass* at a lower degree.

A global analysis of the experimental conditions combining the sensory, hedonic, and physical properties of the presented sound fields allowed the identification of the main drivers of assessors' preferences. The analysis indicated that rooms described by lower RT are preferred. It is however evident that a critical value exists, close to the recommended mean value of 0.4 s [38], above which assessors' preference is degraded.

The current findings support previous studies [5, 6] that the perceived sound field can be significantly altered by the acoustical properties of the reproduction room. Understanding the acoustic influence of these environments on the reproduced sound field will enhance the system's ability to recreate a sonic experience in acoustically-dissimilar enclosures [71], in a more accurate and perceptually relevant way. For example, one could employ directivity control in the loudspeakers, aiming to evoke certain perceptual aspects by controlling the excitation of the room's acoustical field. This desire is evident, as recent advances in spatial audio reproduction over loudspeakers [12] target domestic reproduction environments where the current knowledge is limited.

Notes:

[a] The room is designed to host simulation setups with human occupancy, and meets the requirements for anechoic performance [72] down to 200 Hz, thus, low-frequency compensation was applied between 65-180 Hz [17].

[b] The correlation of an attribute to the factorial solution is given by $R = \sqrt{R^2_{\text{Dimension1}} + R^2_{\text{Dimension2}}}$. A limitation of R$<$ 0.65 to any of the two dimensions allows the analysis for variables that correlate well to the factorial solution, but not necessarily to one dimension only.

[c] as defined by the assessors.

[d] The perceptual attributes were rated consistently and reliably, as hidden repeats were identified, and the factorial analysis depicted these constructs as separate groups of variables.

Acknowledgments

The authors would like to thank Martin Møller, Jukka Pätynen, Sakari Tervo, Claus Vestergaard Skipper, for their support and helpful input. The research leading to these results has received funding from the European Union's Seventh Framework Program under grant agreement no. ITN-GA-2012-316969.

References

[1] L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York) (2004), pp. 1–653.

[2] M. Barron, "Subjective study of British symphony concert halls", Acta Acust. united with Acust. **66**(1), 1–14, (1988).

[3] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles", J. Acoust. Soc. Am. **132**(5), pp. 3148–3161 (2012).

[4] A. Giménez, R. M. Cibrián, S. Cerdá, S. Girón, and T. Zamarreño, "Mismatches between objective parameters and measured perception assessment in room acoustics: A holistic approach", Build. Environ. **74**, 119–131 (2014).

[5] F. E. Toole, *Sound Reproduction: Loudspeakers and Rooms* (Focal Press, Boston) (2008), 1–550.

[6] S. Bech, “Perception of Timbre of Reproduced Sound in Small Rooms: Influence of Room and Loudspeaker Position”, J. Audio Eng. Soc **42**(12), 999–1007 (1994).

[7] S. Bech, “Timbral Aspects of Reproduced Sound in Small Rooms. II”, J. Acoust. Soc. Am. **99**(6), 3539–49 (1996).

[8] S. E. Olive and F. E. Toole, “The Detection of Reflections in Typical Rooms”, J. Audio Eng. Soc. **37**, 539–553 (1989).

[9] S. Bech, “Spatial aspects of reproduced sound in small rooms”, J. Acoust. Soc. Am. **103**(1), 434–45 (1998).

[10] L. William and E. Sean, “Interaction between Loudspeakers and Room Acoustics Influences Loudspeakers Preferences in Multichannel Reproduction”, in *the 123rd Conv. Audio Eng. Soc.*, October, 5-8, 2007, New York, USA, (Audio Engineering Society, New York), Prepr. No 7176.

[11] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, “Perception of Reverberation in Small Rooms : A Literature Study”, in *Proc. 55th Int. Conf. Audio Eng. Soc. Spat. Audio*,August, 27-29, 2014, Helsinki, Finland, (Audio Engineering Society, New York).

[12] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, “Spatial Sound With Loudspeakers and Its Perception : A Review of the Current State”, Proc. IEEE **101**, pp. 1920–1938, (2013), doi:10.1109/JPROC.2013.2264784 .

[13] J. Herre, S. Member, J. Hilpert, A. Kuntz, and J. Plogsties, “MPEG-H 3D Audio — The New Standard for Coding of Immersive Spatial Audio”,IEEE, J. of Selected topics in Sign. Proc., **9**(5), 770–779 (2015).

[14] R. K. Clifton, R. L. Freyman, and J. Meo, “What the precedence effect tells us about room acoustics”, Percept. & Psychophys., **64**(180), 180–188 (2002), doi:10.3758/BF03195784 .

[15] S. Bech, “Perception of Reproduced Sound: Audibility of Individual Reflections in a Complete Sound Field III”, in *the 100th Conv. Audio Eng. Soc.* May, 1–4, 1996, Copenhagen Denmark, (Audio Engineering Society, New York), Prepr. No 4195.

[16] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo, “Concert hall acoustics assessment with individually elicited attributes”, J. Acoust. Soc. Am. **130**, 835–49 (2011).

[17] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S.H. Jensen, “A Rapid Sensory Analysis Method for Perceptual Assessment of Automotive Audio”, J. Audio Eng. Soc **65**(2), (2017), doi:10.1121/1.3607422.

[18] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, method and application* (Wiley) (2006), 1–442.

[19] R. San Martín and M. Arana, “Uncertainties caused by source directivity in room-acoustic investigations”, J. Acoust. Soc. Am. **123**(6), EL.133–8 (2008).

[20] S. E. Olive, P. L. Schuck, S. L. Sally, and M. Bonneville, “The Variability of Loudspeaker Sound Quality Among Four Domestic-Sized Rooms”, in *the 99th Conv. Audio Eng. Soc.*, October, 6–9, 1995, New York, USA, (Audio Engineering Society, New York), Prepr. No 4092.

[21] C. Pike, T. Brookes, and R. Mason, “Auditory Adaptation to Loudspeakers and Listening Room Acoustics”, the 135th Conv. Audio Eng. Soc., New York, USA, October, 17–20, 2013, (Audio Engineering Society, New York), Prepr. No 8971.

[22] C. Pike, R. Mason, and T. Brookes, “The effect of auditory memory on the perception of timbre”, in *the 136th Conv. Audio Eng. Soc.*, April, 26-29, 2014, Berlin, Germany, (Audio Engineering Society, New York), Prepr. No 9028.

[23] M. Sams, R. Hari, J. Rif, and J. Knuutila, “The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence”, J. Cogn. Neurosci. **5**(3), 363–368 (1993).

[24] B. Shinn-Cunningham, “Learning reverberation: Considerations for spatial auditory displays”, in *Proc. 2000 Int. Conf. Audit. Disp.*, April, 2–5, 2000 (Atlanta, USA).

[25] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, “Spatial decomposition method for room impulse responses”, J. Audio Eng. Soc. **61**(1/2), pp. 17–28 (2013).

[26] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, “Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences”, J. Acoust. Soc. Am. **140**(1), pp. 551-562, (2016).

[27] S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki, “Preferences of critical listening environments among sound engineers”, J. Audio Eng. Soc. **62**(5), pp. 300–314 (2014).

[28] N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. van Waterschoot, and S.H. Jensen, “Perceptual aspects of reproduced sound in car cabin acoustics”, J. Ac. Soc. Am. **Accepted**, — (2017).

[29] P. Larsson and D. Västfjäll, “When what you hear is what you see: Presence and auditory-visual integration in virtual environments”, in *Proc. 10th Annu. Int. Work. Presence*, 11–18 (Barcelona, Spain) (2007).

[30] U. Jekosch, *Voice and speech quality perception – Assessement and Evaluation* (Springer-Verlag, Berlin, Germany) (2005), pp. 54–58.

[31] H. T. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices* (Springer, New York) (1999), pp. 1–471.

[32] J. Delarue, B. J. Lawlor, and M. Rogeaux, *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research* (Woodhead Publishing, Elsevier: Cambridge, UK) (2014), pp. 1-555.

[33] S. Choisel and F. Wickelmaier, “Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference”, J. Acoust. Soc. Am. **121**(1), pp. 388–400 (2007).

[34] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, “Elicitation of attributes for the evaluation of audio-on-audio interference”, J. Acoust. Soc. Am. **136**(5), pp. 2630–2641 (2014).

[35] G. Martin and S. Bech, “Attribute Identification and Quantification in Automotive Audio - Part 1: Introduction to the Descriptive Analysis Technique”, in *Prepr. 118th Conv. Audio Eng. Soc.* (Barcelona, Spain) (2005).

[36] V. Dairou and J.-M. Sieffermann, “A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile”, J. Food Sci. **67**, 826–834 (2002), URL `http://dx.doi.org/10.1111/j.1365-2621.2002.tb10685.x`.

[37] J. Delarue and J. M. Sieffermann, “Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products”, Food Qual. Prefer. **15**, 383–392 (2004).

[38] International Electrotechnical Commission, “IEC:60268, Sound System Equipment— Part 13: Listening Tests on Loudspeakers”, IEC:Geneva, Switzerland, (1998).

[39] International Telecommunication Union, “Rec.775-3 Multichannel stereophonic sound system with and without accompanying picture”, ITU-R:Geneva, Switzerland, (2012).

[40] A. Farina, “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique”, in *the 108th Conv. Audio Eng. Soc.*,February 19–20, 2000, Paris,France, (Audio Engineering Society, New York), Prepr. No 5093.

[41] J. Pätynen, S. Tervo, and T. Lokki, “Amplitude panning decreases spectral brightness with concert hall auralizations”, in *55th Int. Conf. Audio Eng. Soc. Spat. Audio* ,August 24–27, 2014, Helsinki, Finland, (Audio Engineering Society, New York).

[42] International Organization for Standardization, “ISO 3382-1: Acoustics - Measurement of room acoustic parameters, Part 1: Performance Spaces”, (2009).

[43] International Organization for Standardization, “ISO 3382-2: Acoustics - Measurement of room acoustic parameters, Part 2: Reverberation time in ordinary rooms”, (2008).

[44] G. Vairetti, N. Kaplanis, E. De Sena, S. H. Jensen, S. Bech, M. Moonen, and T. van Waterschoot, “The Subwoofer Room Impulse Response database (SUBRIR)”, J. Audio Eng. Soc **Submitted** (2017).

[45] C. Díaz and A. Pedrero, “The reverberation time of furnished rooms in dwellings”, Appl. Acoust. **66**, 945–956 (2005).

[46] International Telecommunication Union, “BS.1116-1: Methods For the Subjective Assessment of Small impairments in Audio Systems Including Multichannel Sound System”,ITU-R:Geneva, Switzerland, (1997).

[47] J. Pätynen, S. Tervo, and T. Lokki, “Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses”, J. Acoust. Soc. Am. **133**(2), 842–857 (2013).

[48] Bang & Olufsen, “Music for Archimedes”, NCB: B&O,Struer, Denmark, (1992).

[49] European Broadcasting Union, “Tech 3253 - SQAM: Sound Quality Assessment Material recordings for subjective tests”, 1–13, EBU: Geneva, Switzerland, (2008).

[50] International Organization for Standarization, “ISO:8253-1 Audiometric test methods Part 1:Pure-tone air and bone conduction”, IEC:Geneva, Switzerland, (2010).

[51] International Telecommunication Union, “Rec. BS.1534-2 Method for the subjective assessment of intermediate quality level of audio systems”, ITU-R:Geneva, Switzerland, (2014).

[52] J. Fields, R. De Jong, T. Gjestland, I. Flindell, R. Job, S. Kurra, P. Lercher, M. Vallet, T. Yano, R. Guski, U. Felscher-Suhr, and R. Schumer, "Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation", J. Sound Vib. **242**, 641–679 (2001).

[53] J. Gower, "Generalized procrustes analysis", Psychometrika, **40**(33), (1975),doi: 10.1007/BF02291478 .

[54] J. Pagès, *Multiple Factor Analysis by Example Using R*, (CRC Press, Boca Raton, FL) (2014), pp. 1–247.

[55] S. Le Dien and J. Pagès, "Hierarchical multiple factor analysis: Application to the comparison of sensory profiles", Food Qual. Prefer. **14**, 397–403 (2003).

[56] S. Lê, J. Josse, and F. Husson, "FactoMineR : An R package for multivariate analysis", J. Stat. Softw. **25**, 1–18 (2008).

[57] S. Lê and T. Worch, *Analyzing Sensory Data with R* (CRC Press, Boca Raton, FL) (2015).

[58] A. Lucas, "amap: Another Multidimensional Analysis Package", (2008), URL `http://cran.r-project.org/package1?4amap, Rpackageversion0.8-4`.

[59] T. H. Pedersen and N. Zacharov, "The development of a Sound Wheel for Reproduced Sound", in *138th Conv. Audio Eng. Soc.* (Warsaw, Poland) (2015).

[60] A. Sotiropoulou, "Concert hall acoustic evaluations by ordinary concertgoers: I, Multi-dimensional description of evaluations", Acta Acust. **81**, 10-19 (1995).

[61] A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms", Nature **397**, 517–520 (1999), URL `http://www.nature.com/doifinder/10.1038/17374`.

[62] P. Volk, S. Bech, T.H. Pedersen, and F. Christensen, "Modelling perceptual characteristics of loudspeaker reproduction in a stereo setup", J. Aud. Eng. Soc. **Submitted**, xx–xx (2016).

[63] P. Volk, T.H. Pedersen, S. Bech, and F. Christensen, "Modelling perceptual characteristics of prototype headphones", in *Proc. Intern. Conf of Aud. Eng. Soc on Headphone Technology* (Aalborg, Denmark) (2016).

[64] P. Volk, M. Lavandier, S. Bech, and F. Christensen, “Identifying the dominating perceptual differences in headphone reproduction”, J. Acoust. Soc. Am. **140**(5), 3664–74 (2016).

[65] T. Næs, P. B. Brockhoff, and O. Tomic, *Statistics for Sensory and Consumer Science* (2009), 127-149.

[66] D.A. Berkley,J.B. Allen, “Normal listening in typical rooms: The physical and psychophysical correlates of reverberation”, in *Acoustical Factors Affecting Hearing aid Performance*, 3–14 (1993).

[67] P. Zahorik, “Perceptually relevant parameters for virtual listening simulation of small room acoustics”, J. Acoust. Soc. Am. **126**, 776–91 (2009).

[68] S. Bech, “Timbral aspects of reproduced sound in small rooms. I”, J. Acoust. Soc. Am. **97**, 1717–26 (1995).

[69] S. E. Olive, “A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part I - Listening Test Results”, in *the 116th Conv. Audio Eng. Soc.*, May 8–11, 2004, Berlin, Germany, (Audio Engineering Society, New York), Prepr. No 6113.

[70] B. Blesser and L. Salter, *Spaces Speak Are you Listening* (MIT Press),22-23 (2007).

[71] C. C. J. M. Hak and R. H. C. Wenmaekers, “The impact of sound control room acoustics on the perceived acoustics of a diffuse field recording”, WSEAS Trans. Signal Process. **6**, 175–185 (2010).

[72] International Organization for Standardization, “ISO:3745 - Determination of sound power levels and sound energy levels of noise sources using sound pressure. Precision method for anechoic rooms and hemi-anechoic rooms”, (2003).

Part III

Appendix - Supplementary Papers

Paper E

Spatial Analysis and Synthesis of Car Audio System and Car Cabin Acoustics with a Compact Microphone Array

Sakari Tervo, Jukka Pätynen, Neofytos Kaplanis, Morten Lydolf,
Søren Bech, Tapio Lokki.

The paper was published in
Journal of Audio Engineering Society, Vol 63 (11), pp. 914-925. 2015 .

The layout has been revised.

Abstract

This engineering report describes the application of the Spatial Decomposition Method in a car cabin using a compact microphone array. The proposed method provides objective analyses of the sound field with respect to direction and energy, and enables the synthesis for multi-channel loudspeaker reproduction. Due to the extreme acoustics of a car cabin, a number of recommended steps is presented to improve both objective and perceptual performance of the method. The results suggest that the method can be successfully applied to car audio systems.

1 Introduction

Car audio systems are one of the most listened audio reproduction systems globally, and a few years ago Toole suggested [1, p.41] that most of the listening time is spent in private environments such as cars and home environments. Recently, the increased popularity of mobile devices has probably elongated the total listening time with headphones. Still, cars and their audio systems remain as one of the main environments where listening takes place. For example, a recent survey found that American consumers spent in total 23 % of their listening time in cars [2]. Car audio systems and car cabin acoustics are active research areas. While dummy-head measurements are an often applied approach in the objective and subjective studies [3–12], here we describe an alternative way of obtaining the spatial analysis and synthesis of the car audio systems in a car cabin via compact microphone array measurements.

Analysis and synthesis from microphone array measurements provide more freedom in several aspects when compared to dummy-head measurements. Namely, the spatial sound field can be analyzed more accurately in 3-D, if the applied microphone array sets a 3-D space. In addition, the spatial sound reproduction is not limited to binaural reproduction but loudspeaker reproduction can also be used. A loudspeaker system that surrounds the listener avoids the lack of externalization encountered in many headphone reproduction systems [13]. Thus, a natural externalization allows a more accurate perceptual evaluation of the spatial attributes. Furthermore, loudspeaker reproduction also enables the evaluation of attributes related to the so-called "body impact", which is the physical impact of low frequency reproduction one can experience and is in some cases an important part of the sound design [14, 15].

Binaural measurements with a static dummy-head are commonly applied in spatial sound reproduction. The studies of car cabin acoustics [3–6, 12, 16] use the dummy-head in the driver's position and measure impulse responses from the left and right playback-channels to the left and right artificial ear of the dummy-head, resulting in a binaural room impulse response (BRIR). These BRIRs are then convolved with the appropriate source material for listening

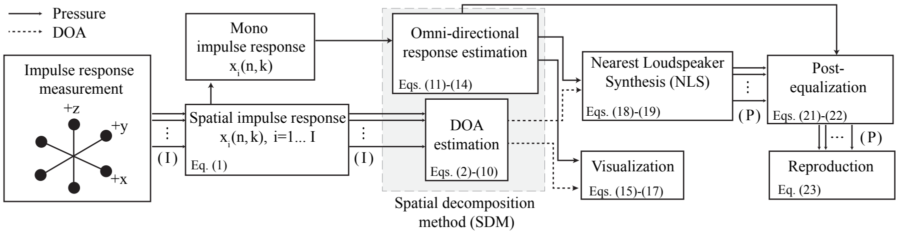


Fig. E.1: Overall processing scheme in the proposed parametric spatial analysis and synthesis.

tests or analyzed in time and frequency to obtain objective metrics. Spatial sound reproduction is obtained via headphone reproduction with appropriate compensation of the headphone responses. Comparisons between different systems in listening tests are easy to implement since instant switching between systems is possible.

Lack of degrees of freedom in static binaural measurements has lead to extending the method to the binaural car scanning (BCS) method [8–11, 17, 18], where a BRIR of the left and right playback-channels are measured over a certain grid of horizontal and lateral angles of the dummy-head. This approach allows natural head movements in the listening test when the head-orientation of the listening test subject is tracked [17]. BCS has improved and gained popularity recently due to the advancements in head tracking hardware and the increase in computational power for multichannel filtering. Nowadays BCS can be considered as one of the main spatial synthesis approaches in car audio research, possibly due to its straightforward processing.

The more indepth analysis of acoustics inside the car requires the use of microphone arrays in the measurements. Microphone arrays are applied in the car audio system and the car cabin acoustics studies to synthesize the acoustics [16], to analyze sound field inside a car cabin [5, 19, 20] and to verify the simulation results of the sound field inside the cabin [21, 22]. A typical setup in these studies includes tens or hundreds of measurements with omni-directional microphones. Another application, where microphone arrays are typically used inside the car cabin is the sound field equalization [23, 24].

Spatial sound analysis and synthesis with the state-of-the-art method in car audio research, BCS, is slow since it requires a large grid of measurements. A measurement with a typical angular resolution of one degree may take up to several hours for a single receiver position. The slowness introduces time-dependency to the measurements, which has to be taken into account in the analysis and synthesis. In addition, with BCS, the head related transfer functions are limited to the ones the dummy head has, which in turn introduces localization, externalization, and timbral errors [13]. Moreover, the sound field can be analyzed in 3-D with the dummy-head measurements if the head is rotated also in the median plane, but the analysis be cumbersome in practice due to the directionality of the dummy head.

In this paper, we propose a spatial sound analysis and synthesis approach for car audio systems, where the acquisition of the measurements is much quicker than in BCS, which allows a loudspeaker-based spatial sound synthesis that does not suffer from the problems that are typical in binaural reproduction with dummy head measurements, and where the 3-D analysis of sound field is simple. The approach is based on the recently introduced parametric spatial sound analysis method, the Spatial Decomposition Method (SDM) [25]. This engineering report is aimed at other researchers and practitioners to provide insight into how to use compact microphone array measurements for analysis and synthesis of car audio systems as an alternative or parallel approach to the dummy-head measurements and binaural reproduction.

2 Methods

SDM analyzes the Direction of Arrival (DOA) in short time windows for each discrete time instant in a spatial room impulse response. Each time moment is assigned the DOA estimate and a pressure value $x_o(n)$ from an omni-directional microphone located ideally in the center of the array. SDM is therefore a parametric method and the spatial room impulse response is described as a set of image-sources or plane waves. Due to the fact that the SDM coefficients are image-sources or plane waves, they are straightforward to synthesize with any of the existing spatial sound reproduction method. Next we review the analysis in SDM assuming one reflection per time window. In addition, the spatial sound reproduction using the nearest loudspeaker synthesis (NLS) is described. For the convenience of the reader, Fig. E.1 shows the processing steps applied in this paper.

2.1 Analysis

A measured wide-band impulse response pressure in microphones $i = 1, 2, \ldots, I$ is presented in the frequency domain as the sum of all acoustic events $r = 1, \ldots, R$

$$x_i(k) = \sum_{r=1}^{R} c_{i,r}(k) h_{i,r}(k) s_r(k) + n_i(k), i = 1, 2, \ldots, I, \tag{E.1}$$

where k is the discrete frequency, $h_{i,r}(k)$ and $s_r(k)$ are the frequency and the source response from the loudspeaker in the direction of the acoustic event r. Moreover, $c_{i,r}(k)$ is the response of the i^{th} microphone in the direction of the acoustic event r, and $n_i(k)$ is a noise component, assumed independent and identically distributed (i.i.d.) for each microphone.

DOA Estimation

In SDM, we assume that a short time window of the room impulse response includes one wide-band reflection, even if this is not true in the late part of a typical impulse response [25]. However, most of the energy in a room impulse response is concentrated in the beginning, where also the echo density is the smallest and the analysis is more correct. Due to these reasons, the perception of the acoustics is also maintained to a large extent, as the beginning of the impulse response has the largest impact on it [26].

Before presenting the details of the proposed method for analysis of car acoustics and audio, we review the theory and the assumptions in SDM. Discussion on the assumptions in Section 3 and more details on the DOA estimation are found in [25].

The wave-front r within the compact microphone array is modeled as a plane wave, and the time delay associated with it in a certain microphone i can be written in the Cartesian coordinate system as

$$t_i = (d_r + \mathbf{m}_i \mathbf{n}_r^T)/c, \tag{E.2}$$

where d_r is the distance from the origin of the wave-front to the center of the microphone array, $\mathbf{m}_i$ is the microphone position w.r.t. the center of the array, $\mathbf{n}_r$ is the normal to the wave-front, c is speed of sound, and $(\cdot)^T$ denotes a transpose of a matrix or a vector. Time difference of arrival (TDOA) of the wave-front in two microphones i and j is written as

$$\tau_{i,j} = t_i - t_j = (\mathbf{m}_i - \mathbf{m}_j)\mathbf{n}_r^T/c. \tag{E.3}$$

We model the measurement of TDOA for each microphone pair as an unbiased process affected by Gaussian i.i.d. noise component w:

$$\hat{\tau}_{i,j} = \tau_{i,j} + w. \tag{E.4}$$

The estimation of the normal $\mathbf{n}_r$ can be presented as Minimum Mean Squared Error (MMSE) problem,

$$\hat{\mathbf{n}}_r = \arg\min_{\mathbf{n}_r} \sum_{\{i,j\}=1}^{M} (\tau_{i,j} - \hat{\tau}_{i,j})^2, \tag{E.5}$$

where the microphone pairs $\{i,j\} = 1, \ldots, M$ are $\{\{1,2\},\{1,3\},\ldots,\{i,j\},\ldots,\{I,I-1\}\}$. For example, with $I = 6$ microphones there are $M = I \times (I-1)/(2) = 15$ different microphone pairs.

The MMSE estimate of DOA in standard Cartesian coordinates is given as [25]

$$\hat{\mathbf{n}}_r \quad = \quad \mathbf{V}^{\dagger}\hat{\tau}, \tag{E.6}$$

where $\mathbf{V} = [\mathbf{m}_1 - \mathbf{m}_2, \mathbf{m}_1 - \mathbf{m}_3, \ldots, \mathbf{m}_I - \mathbf{m}_{I-1}]$ is the difference between pairs of the microphone positions, $(\cdot)^\dagger$ denotes the Moore-Penrose pseudo-inverse, and $\hat{\tau} = [\hat{\tau}_{1,2}, \hat{\tau}_{1,3}, \ldots, \hat{\tau}_{I,I-1}]$ are the TDOA estimates between each microphone pair. The final DOA is obtained by normalizing the estimate to unity as $\hat{\mathbf{n}}_r / \|\hat{\mathbf{n}}_r\|$.

The TDOAs are estimated as the maximum argument of the cross correlation vector

$$\hat{\tau}_{i,j} = \arg\max_{\tau} \left(R_{i,j}(\tau)\right) / f_s, \tag{E.7}$$

where f_s is the sampling frequency, and the cross correlation vector between microphone signal $\mathbf{x}_i$ and $\mathbf{x}_j$ is calculated as the inverse Discrete Fourier Transform (IDFT)

$$R_{i,j}(\tau) = \text{IDFT}\left(\mathbf{x}_i \mathbf{x}_j^*\right)(\tau), \tag{E.8}$$

The discrete Fourier-domain samples of x_i are obtained via the DFT

$$\begin{aligned} \mathbf{x}_i &= [x_i(n,1), \ldots x_i(n,K)] && \text{(E.9)} \\ &= \text{DFT}\left([x_i(n - L/2), \ldots x_i(n + L/2 - 1)]\right)(k), && \text{(E.10)} \end{aligned}$$

n is the time frame, L is the length of the frame, and $k = 1, \ldots, K$ are the frequency bins in the DFT.

This DOA estimation framework is applied in small time windows to the spatial room impulse responses measured in the car with the car audio system. As the DOA estimate is obtained at every discrete time sample n the estimated normal is denoted with $\hat{\mathbf{n}}_n$. Later on in this paper, also standard spherical coordinates, the azimuth angle $\hat{\phi}_n \in (-180, 180]^\circ$ and the elevation angle $\hat{\theta}_n \in (-90, 90]^\circ$ will be used to denote the DOA.

Omni-directional response estimation

SDM assumes that an omni-directional impulse response $x_o(n)$ in the center of the array is available. A typical commercial microphone array rarely features a microphone in the center position, but the sound pressure in the center has to be estimated. Furthermore, real omni-directional microphones are not entirely omni-directional in all audible frequencies and in all directions of incidence. Next, we will present the omni-directional response estimation assuming that there is one plane wave present in the time window, the DOA is known or can be estimated, and that the directional responses of a microphone are available.

An exact solution to obtain the omni-directional flat response from one of the microphones' signals, say, from the i^{th} microphone, is to equalize the directional response in the DOA, as proposed in [25]:

$$\hat{x}_o(n,k) = \frac{x_i(n,k)}{a_i(\mathbf{n}_n, k)}, \forall\, n, k, \tag{E.11}$$

where $a_i(\mathbf{n}_n, k)$ is the microphone response to direction $\mathbf{n}_n$ in frequency k and at time instant n. The microphone response $a_i(\mathbf{n}_n, k)$ can be either measured in anechoic conditions or modeled and $\mathbf{n}_n$ can be estimated as shown in the previous section. The processing in Eq. (E.11) will produce the exact omnidirectional response in the absence of noise in the far-field with the plane wave assumption. Appropriate processing may be required for $a_i(\mathbf{n}_n, k)$ to ensure that $a_i(\mathbf{n}_n, k) \neq 0 \, \forall \, n, k$.

If the directional response $a_i(\mathbf{n}_n, k)$ is unknown, but the DOA Ω_n is available we can apply sub-optimal methods. Instead of the exact equalization filter $1/a_i(\mathbf{n}_n, k)$ we apply some general average equalization. Here, we use the diffuse field (DF) equalization

$$\bar{a}(k) = \int_S \|a_i(\mathbf{n}, k)\| dS, \tag{E.12}$$

where S is the unit sphere. The single microphone response in Eq. (E.11) is then given as

$$\hat{x}_o(n, k) = \frac{x_i(n, k)}{\bar{a}(k)} e^{i2\pi k t_{i,n}}, \forall \, n, k. \tag{E.13}$$

The time delay with the plane wave assumption is given by $t_{i,n} = \mathbf{m}_i \mathbf{n}_n^T / c$, where $\mathbf{m}_i$ is the microphone position, $\mathbf{n}_n$ is the normal of the DOA. The DF equalized response in a single microphone is obtained by setting the time delay to $t_{i,n} = 0$ in Eq. (E.13), i.e.,:

$$\hat{x}_o(n, k) = \frac{x_i(n, k)}{\bar{a}(k)}, \forall \, n, k. \tag{E.14}$$

2.2 Visualization

In this paper, we apply the spatiotemporal visualization technique presented in [27], with the exception of the direction of integration.

The backward integrated Directional Energy Response (DER), which can be considered as an energy histogram w.r.t. the angle ϕ, starting from time moment n_t after the direct sound is calculated as:

$$E(n_t, \phi) = 10 \log_{10} \left\{ \sum_{n=n_0+n_t}^{N} \|x_o(n)\|^2 |\cos(\hat{\theta}_n)| C(n, \phi) \right\}, \tag{E.15}$$

where n_0 is the arrival time of the direct sound and N is the length of the impulse response in samples. The binary decision if a DOA exists at time moment n is given as:

$$C(n, \phi) = [\hat{\phi}_n / \Delta\phi] \Delta\phi \wedge \phi, \tag{E.16}$$

$$\phi = -180^\circ + q \times \Delta\phi, q = 1, 2, \ldots, 360/\Delta\phi^\circ, \tag{E.17}$$

where $\Delta\phi$ is the resolution of the DER, selected by the user, $[\cdot]$ denotes rounding to the nearest integer and $\wedge$ is the logical and operator. In Eq. (E.16), the variable ϕ spans from $-180°$ to $180°$ with $\Delta\phi$ intervals. Moreover, Eq. (E.15), shows DER in the median plane. DER in other planes, transverse and lateral planes, is calculated by changing the coordinate system to the respective domain, as in [27]. As these DERs are presented in selected time instants n_t, a spatiotemporal visualization is achieved, as will be demonstrated in Section 2.2.

2.3 Synthesis

Synthesis of the SDM coefficients is straightforward with any spatial sound synthesis/reproduction method. Originally, we applied vector-base amplitude panning (VBAP) [28] with SDM [25], but in this paper we present the nearest loudspeaker synthesis for the SDM samples. This synthesis approach increases the perceived clarity of the spatial sound synthesis, which is a preferred feature according to our study in [29].

The nearest loudspeaker synthesis (NLS) in a playback system with P loudspeakers for all discrete samples $n = 1, \ldots, N$ is given by

$$\begin{aligned} p &= \arg\min_{s}(\|\hat{\mathbf{n}}_n - \mathbf{n}_s\|), s = 1, 2, \ldots, P && \text{(E.18)} \\ x_p(n,k) &= \hat{x}_o(n,k)e^{-i2\pi k(n-n_p)}, && \text{(E.19)} \end{aligned}$$

where $\hat{\mathbf{n}}_n$ is the estimated DOA, $\mathbf{n}_p$ is the normal vector in the direction of the loudspeaker, s denotes the index of the nearest loudspeaker out of $\mathbf{n}_s$ to $\mathbf{n}_n$, and $n_p = d_p/c$ is the delay from the listening position to the loudspeaker. In several listening rooms, the loudspeakers cannot be arranged to an equal physical distance from the listening position due to the geometry of the room. Thus, the delay n_p compensates for the delay caused by the loudspeaker arrangement, if the loudspeakers are at different distances.

The nearest loudspeaker playback leads to a set of impulse responses $x_p(n), p = 1, \ldots, P$ corresponding to reproduction loudspeakers. NLS has the property that the sum of the loudspeaker response equals the original pressure impulse response $\hat{x}_o(n)$, i.e.

$$\hat{x}_o(n,k) = \sum_{p=1}^{P} x_p(n,k). \qquad \text{(E.20)}$$

Post equalization of the loudspeaker spectra

As discussed above the DOA estimation provides a weighted average of the true DOAs within the analysis window. When the echo density increases, the amount of reflections per time window also increases causing the DOA to change rapidly. Therefore, at each time step, the synthesis loudspeaker may

Table E.1: The transducer system in the studied car and naming. For example, FRT is the Front-Right-Tweeter.

	Left	Center	Right
Front	FLT,FLW,FLM	FCM	FRT,FRW,FRM
Back	BLT,BLW,BLM		BRT,BRW,BRM
Parcel shelf	PLM	PCM, PCS	PRM

The first alphabet is the position in longitudinal axes (F: Front, B: Back, and P, Parcel Shelf) the second alphabet is the position in the vertical axes, (L: Left, C: Center, and R: Right). The last alphabet indicates the type of transducer, (T: Tweeter, M: Midrange, W: Woofer, and S: Subwoofer.)

change in the NLS as the nearest playback loudspeaker s is chosen, as shown in Eq. (E.18). Such abrupt changes introduce excess wide-band transients in the synthesized impulse responses $x_p(n), p = 1, \ldots, P$. As a result, the reproduced loudspeaker power response has the characteristics of a white noise spectrum. This effect will be demonstrated in Section 2.1.

Here the compensation of the spectra is implemented in 50 % overlapping short-time Fourier-transform (STFT) frames such that the spectrum of the omni-directional response matches the sum of the synthesized loudspeaker spectra. For each time window n and frequency band b we apply

$$\hat{x}_p(n,k)_b = \frac{\|x_p(n,k)\|}{\|\hat{x}_o(n,k)\|} e^{-i2\pi k \phi_p(n,k)}, p = 1, \ldots, P, \tag{E.21}$$

where $\phi_p(n,k)$ is the phase of $x_p(n,k)$. The frame size is in this processing three times as long as the lowest wavelength in the frequency band, except for the lowest band it is the average of the longest and shortest wavelength. The final, post-equalized, impulse response for a loudspeaker p is obtained by summing over the individual frequency bands $b = 1, \ldots, B$, as:

$$\hat{x}_p(n) = \text{IDFT}\left(\sum_{b=1}^{B} G_b(k)\hat{x}_p(n,k)_b\right), \tag{E.22}$$

where $G_b(k)$ are the frequency band filters.

Convolution with a source signal

For spatial sound synthesis, each of the loudspeaker impulse responses are convolved with a source signal $s(n)$ as:

$$s_p(n) = \text{IDFT}(\hat{x}_p(n,k)s(n,k)), p = 1, \ldots, P. \tag{E.23}$$

The convolution is often implemented with overlap-save method or overlap-add method if the source signal is long.

3 Spatial Analysis and Synthesis Examples

Experiments were conducted in a four-door sedan-type car with five seats and an interior volume of about 3.0 m^3. The audio system of the car included 17 individual transducers. Five of the transducers were tweeters of which two are acoustical lenses with controlled directivity. In addition, the transducer setup included seven midrange-transducers, four woofers, and one sub-woofer. The overall position of the transducers inside the car is shown in Table 1. The transducer signals were individually amplified with a custom amplifier, and custom software was capable of controlling each 17 transducer channel individually.

A microphone array consisting of six omni-directional microphones was placed in the approximate position of the driver's (person in a car) head. The microphone array was a G.R.A.S. vector intensity probe VI-50 with three coincidental microphone pairs on each axis, separated by 25 mm, shown in Fig. E.1. This microphone array does not have a microphone in the center position. Impulse responses were measured individually from each 17 transducer to the microphones with 5 s long logarithmic sine-sweep at 192 kHz, from 1 Hz to 24 kHz [30]. The environment where the measurement took place was temperature and noise controlled.

The analysis window size L should be as short as possible if wide-band analysis is applied [25]. However, here the analysis is applied to band-limited impulse responses, since the transducers only emit a certain frequency band. Therefore, the analysis window size is selected such that the window length includes at least three periods of the longest wavelength in the frequency band.

3.1 Omni-directional Pressure Estimation and Post Equalization

The omni-directional response estimation is studied with an impulse response measured from Front-Right-Midrange (FRM) transducer to the microphone array. In this experiment, the window size is set to $L = 64$ samples, 99 % overlap is used in the synthesis and the time delay due to DOA is compensated for. The applied DOA in the omni-directional response estimation is obtained from the DOA estimation in Section 1.1.1. The equalization filter $\bar{a}(k)$ is the average of a set of anechoic measurements, as in Section 1.1.2. DF equalization and the estimated response are implemented using Eqs. (E.13) and (E.14).

Figures E.2 and E.3 show the original sound pressure in the top microphone (+z-direction), DF equalized response and the estimated omni-directional response in the time and the frequency domain, respectively. In addition, in Fig. E.3, the DF equalization frequency response $\bar{a}(k)$ is shown. As can be seen from the results, the estimated response has about the same shape as the original signal in both the time and the frequency domain, but has less energy.

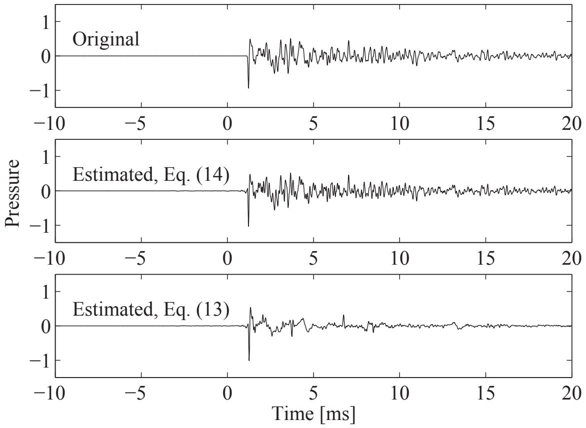


Fig. E.2: Estimation of the omni-directional response for the Front-Right-Midrange transducer in the time domain.

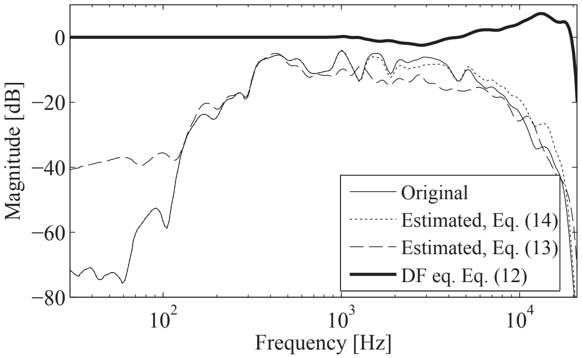


Fig. E.3: Estimation of the omni-directional response for the Front-Right-Midrange transducer in the frequency domain.

The level in the estimated response is lower than in the original response from approximately 1 kHz onwards and higher than in the original response below 100 Hz. On the contrary, DF equalization preserves the approximate shape in the time domain and the frequency domain and increases the level in the high frequencies from 5 kHz onwards with a slight attenuation around 3 kHz (see Fig. E.3). To conclude, DF equalization is a suitable compensation for the microphone spectrum in the present case and is used in the analysis and synthesis throughout the rest of the paper.

The post equalization of the playback loudspeaker spectra is studied with the same impulse response as above. As explained above, the window size is set according to the lowest frequency of each octave band and 50 % overlap is used in the synthesis. The spatial sound synthesis is implemented with NLS and using $P = 24$ loudspeakers, arranged as described for example in [29]. Here, we compare the total magnitude response, i.e., the sum of individual loudspeaker channels' energy

$$E(k) = \sum_{p=1}^{24} \|\hat{x}_p(k)\|^2 \tag{E.24}$$

and the time domain pressure

$$\hat{h}(n) = \sum_{p=1}^{24} \hat{x}_p(n) \tag{E.25}$$

to the corresponding original pressure signal in the frequency and the time domain, which is in this case the DF equalized pressure from the above example. Comparing these results should indicate if the overall spectra and the time domain integrity are maintained. In addition, here we used nine octave band filters in the post equalization of the spectra, and the individual octave band responses are shown in Fig. E.4.

The results of the post equalization of the loudspeaker spectra are shown in Figs. E.4 and E.5 in the frequency and the time domain, respectively. For comparison, the spectra from the NLS of Eq. (E.18) is shown in Fig. E.4. The time domain results for the sum of NLS loudspeaker responses is not presented in Fig. E.4 since it is equal to the original sound pressure, as displayed already in Eq. (E.18). As the results demonstrate, NLS preserves the time domain pressure but distorts the frequency domain spectra, particularly in small enclosures. The distorted spectrum is caused by the rapidly changing DOA estimates. Example of this is shown in Fig. E.6. This effects the NLS so that the loudspeaker in the playback changes also rapidly in consecutive time frames, and causes high frequencies to the spectra whenever the loudspeaker is changed. The post equalization approximately restores the spectrum while maintaining the time domain integrity.

3.2 Visualization Examples

We demonstrate the proposed analysis and visualization with examples from the car measurement, described at the beginning of Section 2. For all the visualizations we use a resolution of $\Delta\phi = 1°$ in the DER. Figure E.6a displays a segment of an impulse response in one microphone measured from one midrange transducer located at the passenger side front door. The analysis window of length $L = 64$ samples (0.3 ms) moves through the entire multi-microphone impulse response from the array. SDM analysis provides direction estimates, and they are shown in Fig. E.6b for azimuth and elevation angles. For clarity,

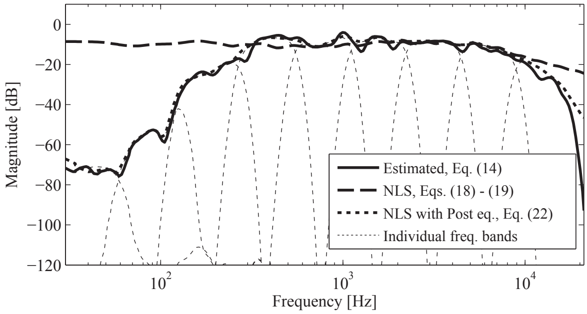


Fig. E.4: The effect of post equalization and the estimated omni-directional response in the frequency domain.

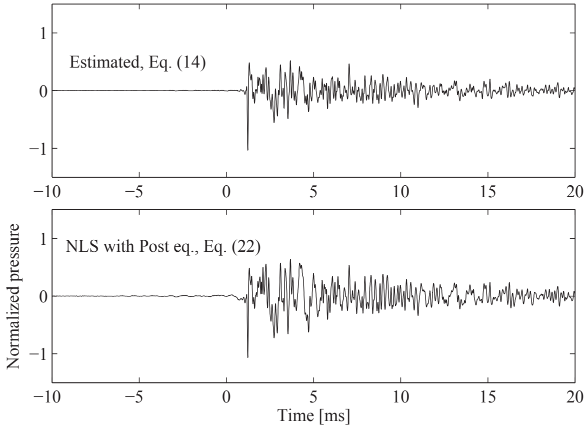


Fig. E.5: The effect of post equalization and the estimated omni-directional response in the time domain.

only every fifth directional value is shown. Figure E.6 demonstrates that during the first peak in the impulse response the direction estimates remain nearly constant up to 0.5 ms, after which the reflections cause the DOA estimation to change along peaks in the pressure. Subsequent reflections appear in the azimuth and elevation as brief stable instants in the direction estimates.

The example in Fig. E.7 shows the spatiotemporal visualizations in the lateral, median, and transverse planes in Fig. E.7a, Fig. E.7b, and Fig. E.7c, respectively. The first time window displays the directional characteristic in the whole impulse response, from the beginning of the direct sound to the end. The next three time windows are backward-integrated and the first two of these discards the direct sound. In other words, they visualize the acoustics of the car interior immediately after the direct sound. The fourth window includes DER

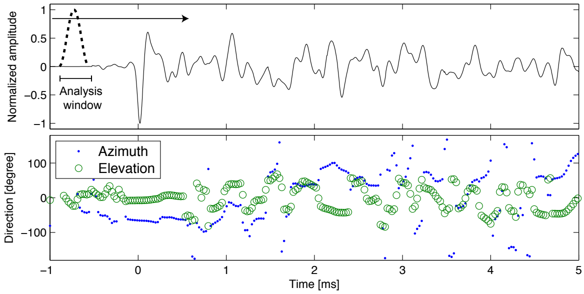


Fig. E.6: Pressure and DOA estimates of the Front-Right-Midrange transducer.

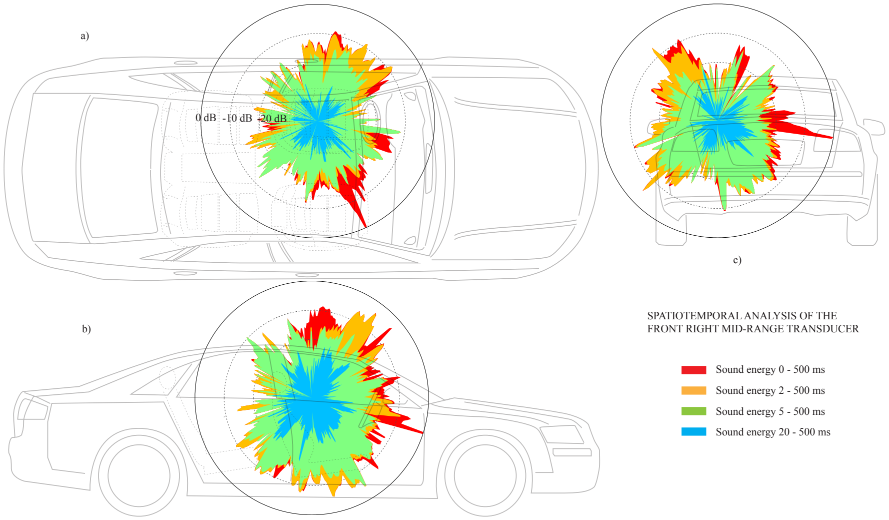


Fig. E.7: Spatiotemporal visualization based on the directional energy response for the Front-Right-Midrange transducer

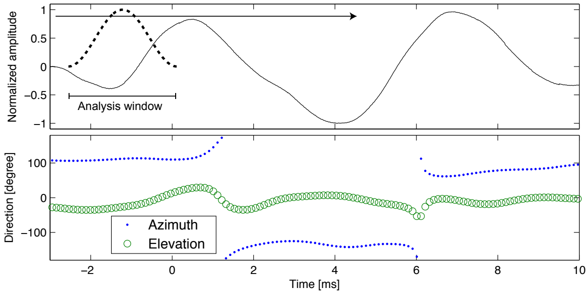


Fig. E.8: Pressure and DOA estimates of the Back-Left-Woofer transducer.

during late decay. We see that the first window already contains reflections from the windshield, driver's side window, and ceiling. With car interior this is reasonable, as 2 ms delay corresponds to approximately 0.7 m longer path of propagation. The largest contributions to the sound field after the direct sound arrive from the areas of the driver's door and the rear seat. The late energy after 20 ms is more uniformly distributed w.r.t. to the early energy.

Second example applies the same approach for a different transducer. Figure E.8 shows the respective analysis for a low-frequency woofer at the left rear door. The wavelength of the response is considerably longer than with the midrange transducer, and the analysis window is adjusted to $L = 512$ samples (2.7 ms), accordingly in order to obtain reasonable TDOA values, and directional estimates. Low audio frequency and small acoustic volume leads to a superposition of the direct and reflected sound. The initial DOA drifts as the reflected sound builds up during the underlying direct sound oscillation.

The superposition of sound waves causes prominent artifacts in the spatiotemporal visualization. The peak in the direct sound points to the correct location in the lateral plane (Fig. E.9a) but the directions for reflections are obscured unlike with the previous example. The cosine weighting in Eq. (E.15) reduces the visualized energy of the actual direct sound in the median plane (Fig. E.9b). The other time windows appear less uniform than in the visualized midrange element, see Fig. E.7. Biased direction estimation due to prominently overlapping reflections is one plausible cause for this result. However, a more plausible explanation is that the low-frequency excitation inside a small interior volume causes standing wave resonance. The direction of the decaying energy suggests a tangential mode in the lateral plane. According to cabin dimensions, the first modes yield resonance frequencies at around 80 Hz (axial, length), 120 Hz (axial, width), and 140 Hz (tangential, length×width). These resonances lie within the transducer frequency range. A waterfall visualization of the spectrogram in Fig. E.10 also suggests resonances at the particular frequency range.

3.3 Synthesis

Informal listening experiments were organized in a listening room in Aalto University. The listening room has 24 loudspeakers arranged in 3-D and surrounding the listening position, as described in Figure E.11. The distance of the loudspeakers to the listener is approximately 1.6 m. The walls and the ceiling of the listening room are treated with absorptive material and there is a carpet on the floor. The wide-band reverberation time of the listening room is about 0.11 s. All 17 transducers were processed using the above described techniques, and three stereo source signals of popular songs were used in the convolution in the auralization.

In the informal listening tests, the authors S.T., J.P., and T.L., compared

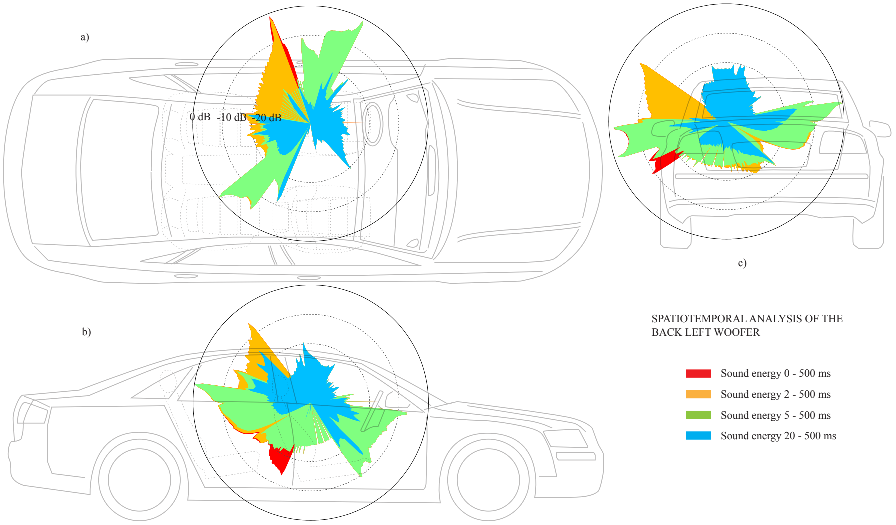


Fig. E.9: Spatiotemporal visualization based on the directional energy response for the Back-Left-Woofer transducer.

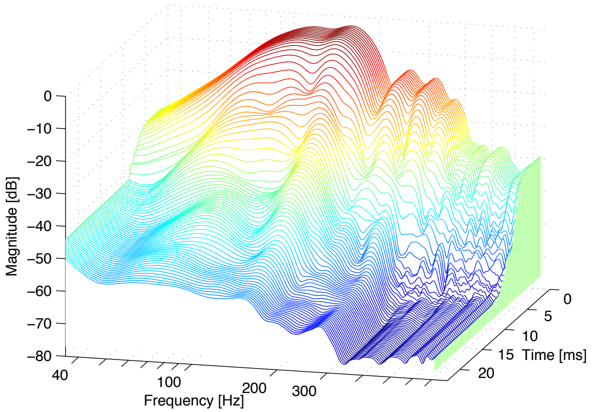
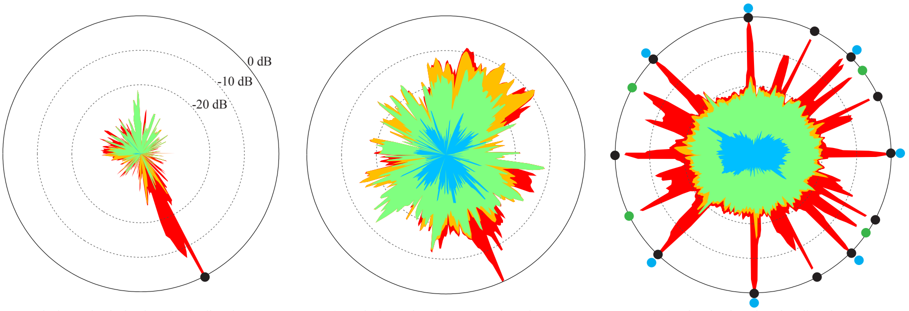


Fig. E.10: Cumulative sound energy, i.e., Schroeder integration, of the Back-Left-Woofer (BLW) response in frequencies [30,800] Hz illustrated as a waterfall visualization. Modes are visible in frequencies from 50 Hz to 200 Hz.

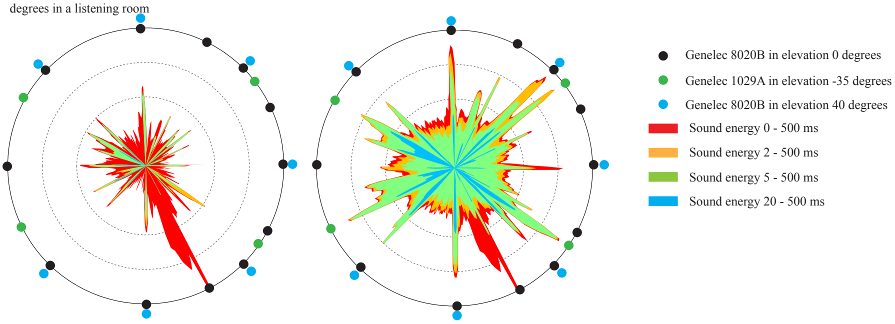
the plausibility of three aspects, spectral fidelity, temporal integrity, and spatial sound image for the cases when the omni-directional estimation and the post equalization were applied and when they were not applied. The most plausible spatial sound reproduction was achieved when using the DF equalization in the omni-directional response estimation and the octave band post equalization for the loudspeaker reproduction. The case when the post equalization was applied



a) Analysis result of a loudspeaker in direction (0,-67.5) degrees in a listening room

c) Analysis results of FRM transducer in a car.

e) Analysis of 24 loudspeakers in a listening room



b) Synthesis of the sound field in a) through 24 real loudspeakers in a listening room

d) Synthesis of the sound field in c) through 24 real loudspeakers in a listening room

Fig. E.11: Spatial analysis and synthesis results in the lateral plane of the so-called *the photocopy of the photocopy tests*. The experiments were conducted in a listening room with 24 loudspeakers. The loudspeaker directly above the listening position in the direction $(90, 0)^\circ$ is not shown in the visualization.

but the omni-directional response was not applied, was perceived to be missing brightness or brilliance. Moreover, the case when the omni-directional response estimation was applied with DF equalization, but there was no post equalization, was perceived too bright, having too much low frequency content, and spatially somewhat vague. Similar spectral bias was observed in all the cases when there was no post equalization.When omni-directional response estimation with DOA and the post equalization were applied, the sound was perceived as more open, spatially sharp, but less bright than with the DF equalization. To conclude, based on the informal listening tests, the post equalization has larger impact on the spectral fidelity than any of the omni-directional response estimation shown here and the spatial cues are preserved, independent of the equalization, as the same DOA estimates are used in all the methods.

3.4 The Photocopy of the Photocopy Test

To investigate the spatial accuracy of the proposed systems two tests are conducted in the above described listening room. In the first test, the microphone array is placed in a listening spot and a spatial room impulse response is measured from a loudspeaker to the array. The measured response is then analyzed with SDM as before in this paper. Then, the SDM coefficients are synthesized using the NLS principle and the proposed post equalization scheme. The achieved 24 impulse response are then convolved with a logarithmic sine sweep and the spatial impulse responses are measured for all the 24 channels. These responses are analyzed again with SDM. The same procedure is repeated for the FRM transducer in the studied car.

Figure E.11 shows spatiotemporal visualizations of the two tests in the lateral plane. As can be seen in Fig. E.11a, the direct sound is accurately localized, the room is highly absorptive and the strongest reflection that the applied loudspeaker produces is less than 20 dB lower than the direct sound. We can see from Fig. E.11b, here the analysis results in Fig. E.11a, are played back from the 24 channels and analyzed again, that the proposed spatial analysis and synthesis maintains the main spatial features in the sound field. Similarly, by comparing Figs. E.11c and E.11d, we observe that the spatial image obtained in the listening room resembles closely the original one. For completeness, Fig. E.11e shows the spatiotemporal analysis of all the 24 loudspeakers in the room.

4 Discussion

As explained above, at some point two reflections overlap in the analysis window. Then, the assumption on the DOA estimation is incorrect and the estimation produces a weighted average direction of the two DOAs of the reflections. The weighting approximately follows the energy of the reflections. This estimation is sub-optimal, as the goal is to estimate the DOA as accurately as possible. However, if the room impulse response is sparse enough, the assumption of one reflection per analysis window can provide meaningful results for both analysis and auralization as shown in [25, 27].

The wave-front within the compact microphone array is modeled as a plane wave. It should be noted that, this does not imply that the assumption would be that the sound wave travels as a plane wave through the entire car cabin. Only that in a very small space, the wave-front may be approximated as a plane wave. As the dimensions of the microphone are small, the curved wavefront, even in the close proximity of the array can be approximated as a plane wave.

The source close to the receiver creates a significant phase and gain mismatch between pressure and particle velocity. This may be a problem for approaches which use the sound intensity measurements, such as Ambisonics. However, for the presented approach this is not a problem, as the propagation model is based on geometrical differences and the estimation is based solely on pressure.

As mentioned above, SDM assumes wide-band reflections in the analysis. This is not generally a problem if the source signal is wide-band, such as a source in the far-field, as shown in [25]. However, in cars, the individual speakers are typically band-limited and in the near-field. Therefore, the post equalization has to be applied which normalizes the spectra such that the sum of the loudspeaker energy responses approximately matches the original omni-directional response and preserves the main features in the time domain. Thus, the processing has to be applied for short time windows to maintain the time domain integrity, and in frequency bands to match the spectra as presented.

5 Conclusions and Future Work

We presented an alternative method for examining the acoustics and sound systems in cars. The method analyzes the DOA with SDM from spatial room impulse responses, measured with a small and compact microphone array for each individual transducer. Since SDM is a parametric approach to spatial impulse response analysis, this allows, for example, that car audio systems can be tuned in a laboratory environment instead of in-situ. In addition, the visualization of the spatial sound can be applied to investigate what is the DOA of sound for a certain transducer.

Future work on this topic includes measurements of a wide variety of car cabins and car audio systems. These measurements will be applied in listening tests implemented in the sensory evaluation framework. The preparations for this listening test are undergoing as the authors are constructing a listening test setup, where the proposed method will be applied to study the differences between different car interior setups. Moreover, the in-depth analysis of the sound field inside the car with the proposed approach is in the future work of the authors.

Acknowledgements

The research leading to these results has received funding from the Academy of Finland project no. [257099] and from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ITN-GA-2012-316969.

References

[1] F.E. Toole. *Sound reproduction: Loudspeakers and rooms.* Taylor & Francis, New York, NY, USA, 2008.

[2] Nielsen Co. MUSIC 360 2014 U.S. study. `http://tinyurl.com/pbdqsf8` Accessed 2015/11/06.

[3] E. Granier, M. Kleiner, B.-I. Dalenbäck, and P. Svensson. Experimental auralization of car audio installations. In *Audio Eng. Soc. Convention 98, Paris, France*, 1995.

[4] S. Squartini, F. Piazza, R. Toppi, A. Lattanzi, E. Ciavattini, M. Bettarelli, F.and Navarri, and W. Lori. Evaluating different vehicle audio environments through a novel software-based system. In *Audio Eng. Soc. Convention 116, Berlin, Germany*, 2004. Paper Number 6003.

[5] M. Kleiner and C. Lindgren. Objective characterization of audio sound fields in automotive spaces. In *Audio Eng. Soc. 15th International Conference, Copenhagen, Denmark*, 1998. Paper Number 15-007.

[6] A. Farina and E. Ugolotti. Subjective comparison of different car audio systems by the auralization technique. In *Audio Eng. Soc. Convention 103, New York, NY, USA*, 1997. Paper Number 4587.

[7] A. Farina and E. Ugolotti. Spatial equalization of sound systems in cars. In *Audio Eng. Soc. 15th International Conference, Copenhagen, Denmark*, 1998. Paper Number 15-009.

[8] S. Bech, W. Ellermeier, J. Ghani, M.-A. Gulbol, and G. Martin. A listening test system for automotive audio-part 2: Initial verification. In *Audio Eng. Soc. Convention 118, Barcelona, Spain*, 2005. Paper Number 6359.

[9] F. Christensen, M. Lydolf, G. Martin, P. Minnaar, B. Pedersen, and W.-K. Song. A listening test system for automotive audio-part 1: System description. In *Audio Eng. Soc. Convention 118, Barcelona, Spain*, 2005. Paper Number 6358.

[10] P. Hegarty, S. Choisel, and S. Bech. A listening test system for automotive audio–part 3: Comparison of attribute ratings made in a vehicle with those made using an auralization system. In *Audio Eng. Soc. Convention 123, New York, NY, USA*, 2007. Paper Number 7224.

[11] S. Choisel, P. Hegarty, F. Christensen, B. Pedersen, W. Ellermeier, J. Ghani, and W. Song. A listening test system for automotive audio-part 4: Comparison of attribute ratings made by expert and non-expert listeners. In *Audio Eng. Soc. Convention 123, New York, NY, USA*, 2007. Paper Number 7225.

[12] F. von Tuerckheim and T. Münch. Automated sound optimization of car audio systems using binaural measurements and parametric IIR filters. In *Audio Eng. Soc. Convention 137, Los Angeles, CA, USA*, 2014. Paper Number 9166.

[13] D.R. Begault, E.M. Wenzel, and M.R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10):904–916, 2001.

[14] S. Bech. Requirements for low-frequency sound reproduction, part i: The audibility of changes in passband amplitude ripple and lower system cutoff frequency and slope. *J. Audio Eng. Soc.*, 50(7/8):564–580, 2002.

[15] W.L. Martens, H. Sakanashi, W. Woszczyk, and S.E. Olive. Whole-body vibration associated with low-frequency audio reproduction influences preferred equalization. In *Audio Eng. Soc. Conference: 36th International Conference: Automotive Audio*, 2009.

[16] A. Farina and E. Ugolotti. Subjective comparison between stereo dipole and 3D ambisonic surround systems for automotive applications. In *Audio Eng. Soc. 16th International Conference*, 1999. Paper Number 16-048.

[17] P. Mackensen, U. Felderhof, G. Theile, U. Horbach, and R. Pellegrini. Binaural room scanning - A new tool for acoustic and psychoacoustic research. *J. Acoust. Soc. Am.*, 105(2):1343–1344, 1999.

[18] S.E. Olive and T. Welti. Validation of the binaural car scanning system for subjective evaluation of automotive audio systems. In *Audio Eng. Soc. 36th International Conference, Dearborn, Michigan, USA*, pages 2–4, 2009.

[19] M. Binelli, A. Venturi, A. Amendola, and A. Farina. Experimental analysis of spatial properties of the sound field inside a car employing a spherical microphone array. In *Audio Eng. Soc. Convention 130, London, UK*, 2011. Paper Number 8338.

[20] J. Nowak and M. Strauß. Sound field reproduction analysis in a car cabin based on microphone array measurements. In *Audio Eng. Soc. 48th International Conference, Munich, Germany*, 2012. Paper Number 6-3.

[21] M. Strauss, J. Nowak, and D. de Vries. Approach to sound field analysis and simulation inside a car cabin. In *Audio Eng. Soc. 36th International Conference, Dearborn, Michigan, USA*, 2009. Paper Number 10.

[22] A. Celestinos, M. Olsen, M.B. Møller, and M. Lydolf. Car interior simulation model for low frequencies using the finite difference time domain method. In *Audio Eng. Soc. 48th International Conference, Munich, Germany*, 2012. Paper Number:6-4.

[23] S. Cecchi, L. Palestini, P. Peretti, F. Piazza, and A. Carini. Multipoint equalization of digital car audio systems. In *Proceedings of 6th International Symposium on Image and Signal Processing and Analysis, 2009*, pages 650–655, 2009.

[24] S. Cecchi, L. Palestini, P. Peretti, L. Romoli, F. Piazza, and A. Carini. Evaluation of a multipoint equalization system based on impulse response prototype extraction. *J. Audio Eng. Soc.*, 59(3):110–123, 2011.

[25] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki. Spatial decomposition method for room impulse responses. *J. Audio Eng. Soc.*, 61(1/2):16–27, January/February 2013.

[26] A. Haapaniemi and T. Lokki. Identifying concert halls from source presence vs room presence. *J. Acoust. Soc. Am.*, 135(6):EL311–EL317, June 2014.

[27] J. Pätynen, S. Tervo, and T. Lokki. Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses. *J. Acoust. Soc. Am.*, 133(2):842–857, February 2013.

[28] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.*, 45(6):456–466, 1997.

[29] J. Pätynen, S. Tervo, and T. Lokki. Amplitude panning decreases spectral brightness with concert hall auralizations. In *Proc. 55th Audio Eng. Soc. conference*, Helsinki, Finland, Aug. 27-29 2014. Paper no. 49.

[30] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Eng. Soc. Convention 108, Paris, France*, 2000. Paper Number: 5093.

Paper F

A Method for Perceptual Assessment of Automotive Audio Systems and Cabin Acoustics

Neofytos Kaplanis, Søren Bech, Sakari Tervo, Jukka Pätynen,
Tapio Lokki, Toon van Waterschoot, Søren Holdt Jensen.

The paper has been presented at the
60th International Conference of Audio Engineering Society on Spatial Audio,
Helsinki, Finland, August 2014.

2014, AES
The layout has been revised.

Abstract

This paper reports the design and implementation of a method to perceptually assess the acoustical properties of a car cabin and the subsequent sound reproduction properties of automotive audio systems. Here, we combine Spatial Decomposition Method and Rapid Sensory Analysis techniques. The former allows instant comparisons between auralized measured Vehicle Impulse Responses (VIR) over loudspeakers, avoiding headphone-related shortcomings, while rapid sensory analysis overcomes time-consuming product profiling and language-specific problems, commonly found in the evaluation of audio material. The proposed method is described in terms of capturing, analyzing and reproducing the sound field. A brief overview of the experimental procedure is presented as well as preliminary results of a pilot experiment.

1 Introduction

Over the last decades the automotive industry has been focusing in identifying and improving the major factors that influence the sensory experience of the vehicles. As a consequence, the study of sound quality in automotive audio systems has been brought into the limelight.

The highly complex acoustical environment of a car cabin minimizes the effectiveness of standard objective measures, lacking robustness, repeatability and perceptual relevance. This has lead to the use of human auditory perception as a the major instrument in car audio evaluation, both during development as well as aftermarket benchmark processes [1–9]. Aiming towards high-quality reproduction, car audio manufacturers normally employ listening tests to identify and characterize the physical alterations of this type of sound fields. These alterations relate to electroacoustic properties of the transducers, their placement and performance, cabin's properties, as well as signal processing algorithms (i.e. upmixing & sound tuning).

Assessing sound is ultimately a process that requires merging its *perceptual* and *physical* characteristics in a common space. In acoustics, understanding the perceptual effects of the physical properties of a space would enable a better understanding of its acoustical qualities and stipulate perceptually relevant ways to compensate for the subsequent degradation.

The physical characterization of sound fields has mainly focused in performance and typical sound reproduction spaces. Later, several objective metrics have been established and standardized [10, 11]. Numerous studies have also investigated the perceptual constructs of these acoustical fields, aiming to identify the corresponding sensory descriptors and quantify the human sensations. A recent in-depth review of these studies [12] has revealed that these perceptual constructs could be domain specific, as the physical characteristics of the room, complexity of the stimuli, and listener's expectations may affect our perceptual and cognitive processes.

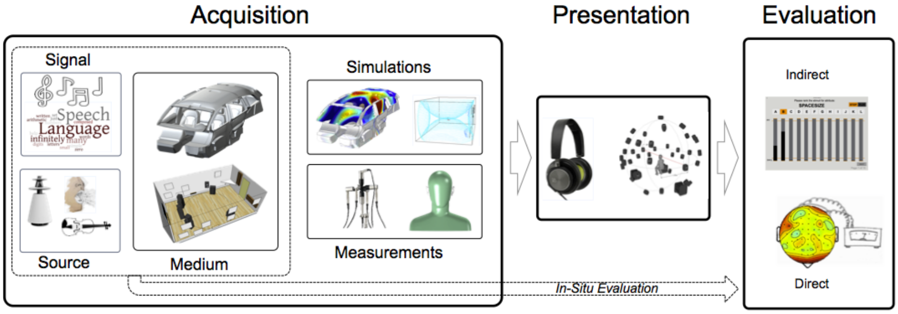


Fig. F.1: Basic principles undertaken in perceptual audio evaluation.

The experimental procedure employed in perceptual evaluation of audio material is normally driven by the contextual factors, given limitations and application of research, which typically translates into a domain-specific standardization. Most experimental methods and recommendations are typically based on a three-step process: *acquisition*, *presentation*, and the *evaluation* of a stimuli set. A graphical summary is given in Fig.F.1. Based on these principles, several standardization schemes have been applied in audio evaluation in several domains. For example a recommendation for experimental procedures for small degradations in audio signals [13] and good practices for evaluation of audio in network transmission protocols [14]. Such a standard is not available for car audio evaluation [3].

Several experimental procedures have been exploited in automotive audio evaluation, which can be classified into two categories: (1) *in-situ* (2) Laboratory-based evaluation. It is logical to assume that *in-situ* evaluation provides the highest *ecological validity* possible. It is therefore not a surprise that such evaluations formed the industry standard. However, exposing assessors in such settings introduces strong biases caused by non-acoustical factors (i.e. size, price, brand). Later, Shively [6] proposed a form of a blind *in-situ* evaluation. Nonetheless, *in-situ* methods inherently include long test-to-test periods between the different *Device Under Test* (DUT) which is known to affect perception thus, its judgment. It is therefore likely that the experimental results are influenced in an uncontrolled manner; i.e. effects of the assessor's auditory memory, mood, and expectation [15].

These limitations were the driving force into developing laboratory-based evaluations methodologies. The most commonly used method incorporates *dummy-head* measurements at the driver's position which are then used to synthesize the binaural field and present the auralization over headphones. A process known as *Binaural Car Scanning* (BCS) [4, 16]. These approaches form a valuable tool in car audio as they overcome practical complications of *in-situ* evaluations, in addition to allowing rapid comparisons between car cabins or audio systems - something not possible previously.

It is however realized that headphones reproduction based on dummy-head measurements result to lack of *externalization* - known as *in-the-head* perception of sound - not accurate timbral reproduction at low frequencies, as well as lack of 'whole-body' vibrations - a sensation known to affect listener preferences in binaural reproduction [17, 18]. In practice, such methods are also time consuming as obtaining *Impulse Responses* (IRs) with a dummy-head requires hundreds of measurements per system.

1.1 Motivation

Perceptual evaluation of acoustic properties in rooms and consequently in car cabins, is highly influenced by the methodology employed, which is typically limited to the specific context factors. A trade-off between the requirements of high *ecological validity* and direct, single-dimensional *variable control* is apparent, imposing application-specific approaches.

In car cabins, the complex sound field requires stimuli acquisition in the form of measurements, or *in-situ* evaluation under real conditions. When assessing *in-situ* a number of non-auditory features and expectations introduce more restrictions. Yet, the benefit of assessing several automotive audio systems and cars in a comparative method seems superior to the uncontrolled auditory memory [19] assumptions during *in-situ* testing.

In this paper, we present an alternative approach towards capturing, presenting, and evaluating automotive sound. The method is based on *Spatial Decomposition Method* (SDM) [20], that has successfully been applied for perceptual evaluation and spatial analysis in both concert halls [21] and critical listening environments i.e. studios [22]. The SDM is a spatial analysis and synthesis scheme, where the obtained IRs can be analyzed parametrically as a pressure and direction of arrival values. Therefore the sound field is decomposed in terms of direction and time of arrival.

Capturing and analyzing the sound field using SDM in cars may lead to two major advantages over the previously discussed methods. First, the spatio-temporal analysis provided by the method may enable better understanding of the behavior of the sound field in the cabin. The method includes additional physical quantities and visualization capabilities, that could be used when the spatial attributes are in question. Second, it allows a variety of reproduction protocols (e.g. *Vector Based Amplitude Panning* (VBAP), *High Order Ambisonics* (HOA)). Thus, a loudspeaker reproduction system could be employed, overcoming the issues of headphone-based playback, such as the lack of 'whole-body' vibration and not externalized sound.

2 Experimental Method

This section describes the proposed audio evaluation methodology for capturing, analyzing, synthesizing, and presenting automotive sound. Here, the reproduction of the auralized stimuli is conducted over loudspeakers in an anechoic chamber and the assessment follows *Sensory Analysis* (SA) procedures, as the optimal combination for this study. It is noted that the presentation and evaluation method could be altered based on the research objectives and the related practical implications. For example presentation on headphones is still possible, as well as following standardized audio evaluation methods as Basic Audio Quality [13], while the spatio-temporal analysis of the measured field will still be available to the researcher.

2.1 Acquisition - In-situ Car Measurements

In order to obtain the acoustic characteristics of a sound reproduction system in a car cabin, *in-situ* recordings of a four-door sedan were performed. The car was equipped with 17 band-limited transducers (5 tweeters, 7 mid-range transducers, 4 woofers, 1 subwoofer) and a custom multichannel amplifier. The audio system included an experimental tuning by a tonmeister, thus the feed to the individual transducers was post-processed (i.e compensation delays, equalization) to represent a typical production car, equipped with a high-end system.

An open spherical microphone array (G.R.A.S VI-50) comprising of two coincidental microphones on each axis, separated by 25mm, was positioned at the driver's seat, at the average seating position [23]. The microphone probe was aligned to match the position of a dummy-head seating in the car - the center point of the head and ears' height. The distance between the microphone array and the headrest was set to 15cm.

The IRs were measured in a way that the (electrical) input to the amplifier was captured by the electrical output of the microphones in the cabin, including any signal processing in the signal path of the audio system. These measurements will be referred to as *Vehicle Impulse Responses* (VIRs).

The VIRs were measured for each transducer, using a 5s logarithmic sine-sweep (1Hz-24kHz) method [24] at 192kHz using an RME UCX multichannel sound interface. The measurement system was calibrated with 5s pink-noise to produce 82dB C-weighted sound pressure level in the car cabin as measured at the forward facing microphone of the array, with system's default settings. The electrical output of the measurement system was kept constant for all drivers. The car measurements were conducted in a temperature and noise regulated garage at Bang & Olufsen's premises.

2.2 Spatial Analysis and Synthesis of VIRs

The spatial analysis and synthesis of the car's sound field was implemented based on SDM [20]. SDM divides the sound field into spatially discrete elements of a preset analysis window. SDM assumes a wide-band source i.e. a typical full-range loudspeaker, therefore it is recommended to use as short window as possible [20, 25]. In this experiment the captured VIRs were band-limited, due to the type/size of each transducer in the cabin. Hence it was possible to implement a custom window-length (L) settings based on the properties of the transducers - at least three periods of the shortest wavelength in the frequency band analyzed. This allows a more accurate spatial decomposition of the sound field - an important advantage when analyzing such complex sound fields as in car cabins. The analysis was performed on the captured VIRs with no modifications as described in 2.1.

In a recent study [25], the spatial analysis and synthesis of a car audio system using SDM is described in detail including experiments conducted to evaluate the undertaken methodology. In that report, experiments have shown that a post-equalization of the loudspeaker spectra should be employed when SDM is used in car cabins. The high echo density found in a car cabin may introduce abrupt changes of the calculated *Direction Of Arrival* (DOA). In consequence excess wide-band transients may be introduced in the synthesized *Impulse Response* (IR). A detailed of the spatial analysis and synthesis procedures followed in the current paper, is given in [25].

2.3 Reproduction Protocol

SDM provides a spatial analysis and signal encoding for a given set of VIRs allowing auralization of the sound field using a given Spatial Decoding and Reproduction scheme over loudspeakers, or Binaural Synthesis based on anechoic *Head-Related Transfer Function* (HRTF). In this paper, the synthesis of the SDM-encoded spatial IRs was implemented using the *Nearest Neighbor* (NN) loudspeaker, similar to [22]. Although reproduction with SDM was first employed using VBAP [26], the synthesis of the soundfield using direct-feeds of SDM-samples to the nearest-loudspeakers with respect to the direction parameters of that sample was found to provide more natural sound without reducing the perceived *brightness* [27].

2.4 Reproduction Setup

In order to provide as close spatial reproduction as possible when new type of environments is to be auralized with SDM, accurate spatial analysis of the soundfield should be performed to identify the placement of the loudspeakers in the reproduction system. As shown previously [25] the directional energy responses of a single mid-range loudspeaker in a car-cabin includes reflections

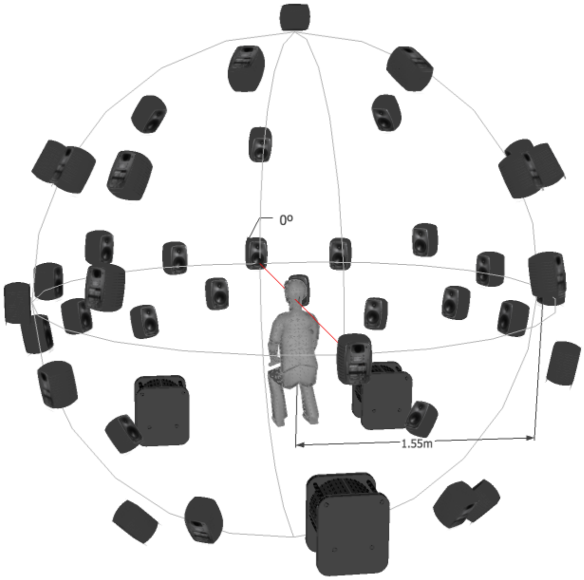


Fig. F.2: The loudspeaker are placed in a spherical orientation (r=1.55m). Loudspeakers are mirrored-placed anti-clockwise in (AZ,EL): 0,0; 11,-10; 22,0; 32.5,-15; 45,5; 55,-10; 65,0; 75,-10; 90,0; 120,-10; 135,10; 0,30; 40,40; 90,30; 115,30; 135,30; 150,55; 0,90; 55,-40; 120,-45; 150,-35.

within very short time intervals i.e. in the case of a Front Right Mid-range that was 0.5 ms as measured at the (left) driver position. Thus, individual *Vehicle Impulse Response* (VIR) analysis was employed to ensure that the direct sound as well as reflections from the cabin's surfaces are preserved, in the best possible way, during the reproduction phase. A systematic approach was followed, using objective and perceptual evaluations. The spatio-temporal energy distribution in time intervals [28] of each measured VIR was combined with the corresponding weighted energy error estimation. This error term, is inherently caused by fitting SDM-samples to the NN loudspeakers instead of their absolute position given in the analysis. The perceptual assessment focused on three perceptual constructs, *spectral fidelity*, *temporal integrity* and *accuracy of spatial representation* for two auralization sets: a full car audio system, and single transducers. Attention was also given to the electroacoustic properties of the loudspeakers and the spatial acuity of the human hearing system, ensuring high level of detail in the frontal plane, whilst maintaining the perceptual qualities from all directions. During the design of the loudspeaker setup, it was noted that a second layer of loudspeakers at lower elevation (-10 °) in the frontal plane (-70°- +70°) was necessary. In fact, the direct sound will only arrive from lower elevations in a car cabin due to the lack of transducers at ear-height at this plane.

For the final auralizations a 40.3 loudspeaker system was specified (see Fig. F.2). The setup comprises of 40 full-range speakers and 3 subwoofers. Each loudspeaker was calibrated with 5s pink-noise *in-situ* at 80.5 ±0.5 dBC$_{RMS}$ measured with a single omni-directional microphone at the listening position. The magnitude response of the loudspeakers was also confirmed to lie primarily within ±1.5dB, including a low-frequency compensation (<180Hz) at the listening position.

2.5 Reproduction Signal Flow

The loudspeakers are driven by a laptop computer over a MADI interface (RME MADIXT). Three D/A converters, 2xRME M16 and 1 x ADI-8 are used to distribute the signals to the individual loudspeakers in the reproduction setup, at 48kHz sample rate. This setup was also used during the calibration procedure including an additional A/D converter (RME Micstacy). The system is self-compensating for internal AD-DA conversion, and has been verified to deliver sample-accurate response at all channels, an overall delay of 18 samples. The ADI-8 introduces an additional 10 sample delay due to MADI-ADAT conversion. Moreover, the inherent inconsistencies in the physical placement of the loudspeakers introduce different time of arrivals at the listening position. Thus, the acoustic delay between each loudspeaker and a microphone at the listening position was measured and the reproduction channels were individually compensated.

2.6 Reproduction Environment

When assessing spatial audio over loudspeakers, it is necessary to decrease the acoustic influence of the reproduction room, on the reproduced soundfield that we intend to evaluate [29], as it is known to be perceptible by listeners [30, 31]. This is normally achieved by ensuring that the reproduction room is characterized by a lower reverberation time compared to the room that is being reproduced via the system. Due to the nature of the sound-field in a car-cabin and the very short reverberation time [32], the setup was installed in the anechoic chamber (B5-104) located at Aalborg University. The chamber is designed and constructed to host simulation setups with human occupancy, and it is treated with absorption wedges that are 0.4m long. Its free inner dimensions are 5.0 x 4.5 x 4.0m. The chamber meets the requirements for anechoic performance [33] down to 200Hz. The physical setup was covered with absorption material to eliminate any reflections from the structural installation.

3 Perceptual Evaluation Methods

Due to the complex and multidimensional properties of audio, most audio perceptual evaluation methodologies require a set of verbal descriptors known as *perceptual attributes*, so that human assessors are able to epitomize and appropriately quantify their sensations for the given set of stimuli. The experimenter may employ an *attribute elicitation* methodology, or define a set of attributes that has been known to form the perceptual space for the given stimuli set [29].

Sensory Analysis (SA) methodologies found in food and wine industry [34, 35] have been instrumental in decoding complex perceptual constructs and hedonic responses of human assessors. Such methodologies hold key importance in multidimensional products such as audio. Over the decade several approaches have been successfully applied in concert halls [21, 36], spatial audio reproduction through loudspeakers [37–41] and headphones [42], hearing aids [43], and active noise cancellation [44].

Descriptive Analysis (DA) is known to be the most sophisticated tool [34] in SA allowing the experimenter to extract information normally hidden behind hedonic and affective judgments, or linguistic inaccuracies leading to a more detailed investigation. It is however known that conventional descriptive SA require extensive training per product, as well as multiple sessions (5-6) per assessor [34, 45]. The time restrictions within the automotive environment and the requirement of product-specific training of DA may not suit well automotive audio. In a need for less time-demanding methods, *rapid sensory profiling* techniques have been proposed recently (see [15] for review).

The most closely related rapid method to conventional profiling is *Flash Profile* (FP) [15, 46]. FP highly emphasises on rapidity aiming to provide a perceptual ‘snapshot’ of the product’s properties as perceived by the assessors and a relative ranking of each product [15, 45]. The biggest advantage over other non-verbal based methods is that it is based on quantitative description, allowing statistical analysis to create a common space i.e. using *Generalized Procrustes Analysis*. The FP only requires 4-5 expert assessors who complete the process within a single 1-3 hours session.

In this study the experimental procedure follows FP recommendations to assesses its applicability in this context, as at the authors knowledge, FP has never been used in audio evaluation before.

4 Pilot Experiment

The sections above described a novel method for perceptual assessment of automotive audio. For reasons of completion, a brief description of a pilot study using the proposed method is included. The data presented here is a part of a formal evaluation that is currently under investigation, and should be interpreted with care. The study's objective was to quantify the perceived differences imposed by altering the acoustics of a car's cabin. In the example below the perceptual effect of the *Front windows* and the effect of the *Equalization* (EQ) (tuning) by an expert tonmeister is in question. The conditions used in the assessment are summarized in table F.1. A short description of the experimental procedure is given below.

4.1 Materials & Apparatus

For this experiment three *in-situ* sets of car measurements were used. Each set included VIRs of all seventeen individually measured transducers, as described in section 2.1. Each VIR was analyzed using SDM. Music material (*Armin van Buuren feat. Ana Criado - I'll Listen*) was then convolved separately with the corresponding 40.3-channel SDM responses. The playback was based on multichannel 24bit PCM sampled at 48kHz.

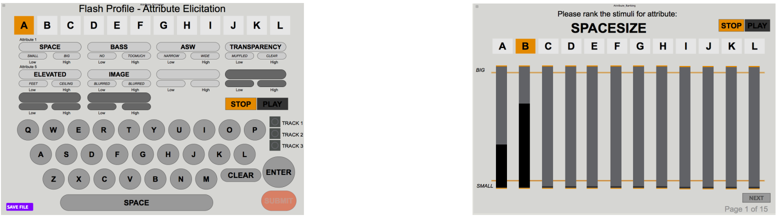
The assessor was given a touch-screen wireless tablet (iPad 2) controlling MAX 7 via MIRA on a Macbook Pro. A custom patch controlled the multichannel audio files and data collection. The interface of the patch was similar to the one found in MUSHRA [47]. The reproduction room, setup and signal chain and calibration measurements were identical to the aforementioned settings in section 2. The assessor was seated on a predetermined location located at center of the spherical array. The height was altered to match the acoustic center of the loudspeakers at 0° elevation using leveling laser. The whole experiment was conducted in complete dark conditions, and the assessor was not aware of the room, loudspeaker setup and the content of the stimuli as suggested by [15, p99].

No	Condition
1	No Equalization (inc. delays)
2	Normal Condition - Tuned
3	Front Windows Open - Tuned

Table F.1: Summary of conditions used in the experiment.

4.2 Experimental Procedure

The experiment comprised of two parts: 1) the *elicitation* of attributes and 2) the *ranking* of the elicited attributes. The assessor was first introduced to the FP methodology, focusing on the correct elicitation of non-hedonic, singular and rank-able descriptors that do not exhibit redundancy [29, 34].



(a) Verbal Elicitation GUI

(b) Attribute Ranking GUI

Fig. F.3: User interface during the two phases of the experiment. Letter buttons included different acoustical conditions of the car cabin. The interface was touch controlled on a tablet.

4.3 Experimental Design - Elicitation

The assessor was asked to provide a non-limited number of verbal descriptors that capture the perceptual characteristics of the whole stimuli set presented. The stimuli were available on a single screen and the assessor was able to play any stimulus at any point, using all tracks provided[1]. The order of presentation was randomized for each assessor. The procedure was self-controlled and self-paced; the GUI is shown in Fig.F.3(a).

4.4 Experimental Design - Ranking

During the ranking phase the assessor was asked to rank the stimuli based on the perceived intensity differences of the given attribute. Each given perceptual attribute formed a block of three trials (one trial per program material). The order of presentation of the stimuli as well as the program material was randomized on each trial. The graphical interface is show in Fig.F.3(b).

4.5 Results

The results of this pilot investigation are given below for a single assessor. The assessor was an expert sound engineer with experience >10 years in developing, evaluating and tuning car audio systems as well as SA methodologies. The assessor is considered as a product expert.

The initial results presented here show that the system is capable of eliciting the expected responses based on the changes in the sound field. Figure F.4 shows that although the 'image focus' is preserved between *Normal Conditions* and *Windows Open*, the perceived 'Width' differs marginally. The perceived 'transparency' seems to decrease when the windows are open, compared to the other conditions presented here. Moreover, perceived 'Distance' shows no

---

[1] This paper only reports one track and three conditions for reasons of simplicity.

difference between the two EQ settings whilst it was ranked higher for *Windows Open*, as one would expect due to lack of side reflections.

These results indicate that the assessor perceived the physical changes in an expected way, which come in close agreement with previous elicitation studies in automotive audio [1] and spatial audio reproduction [12]. It should be noted that the data presented here is an illustrative set and should not be used to conclude findings due to the limited contextual factors. An in-depth analysis will be presented in future work.

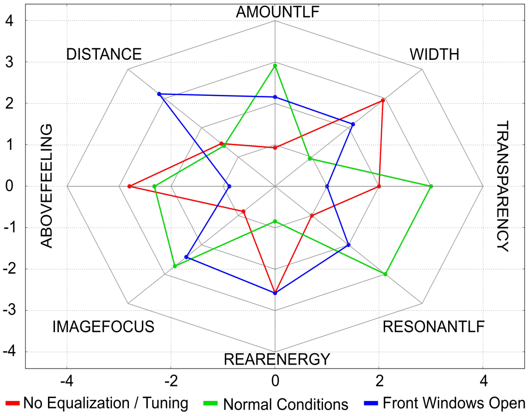


Fig. F.4: Spider plot depicting the responses given by the subject. The values are standardized (M=0). Note that the attributes given here were elicited using a larger set of stimuli.

5 Discussion

This paper described a method to capture, analyze, reproduce and evaluate sound originating from car cabins. The proposed method maintains the benefits of perceptual assessment in the laboratory via auralization similar to BCS. Thus, the method allows rapid and comparative assessment of different sound fields in the lab, in double-blind settings, overcoming issues related to non-auditory feedback (i.e. brand) during *in-situ* evaluation, and the need for a car prototype systems for long periods during development phases. The system also allows a choice of reproduction schemes (i.e. Binaural Synthesis, Wave-Field Synthesis, VBAP). In this study a loudspeaker-based reproduction was followed in an anechoic chamber, ensuring a natural externalization of the reproduced sound field. In addition it includes a full-range reproduction system (23Hz-20kHz) to satisfy low frequency response similar to the car cabin and natural 'whole-body' vibrations, with minimal effects of the reproduction room. Finally, following the proposed method the required time for stimuli acquisition is significantly reduced as capturing the VIRs require less time compared to BCS.

Christensen et al. [4] suggested that providing real driving simulations i.e. Steering Wheel, may enhance the realism of the evaluation. In this study, we aim to remove any visual influence which is known to affect perception, thus, the behavioral outcome (p96 in [15]), an issue raised when performing car audio evaluation.

In the paper Sensory Analysis is also discussed as a tool for audio evaluation. Flash Profile aims for rapid profiling of a product within a single session whilst the assessor can use individual vocabulary. The downside is that the experimenter cannot argue on the semantic meaning of the descriptors. However performing statistical analysis in multiple assessors the perceptual constructs could be decomposed and provide better understanding of the soundfield when merged with the physical metrics within the stimuli.

One should note that the SDM provides a faithful and plausible acoustical representation, however, as any spatial reproduction method to date, it has certain limitations. As it was shown recently [25], a post-equalization of the analyzed response is needed when this is applied to cars due to high echo density. Moreover, the complex geometry of a car cabin and the extreme acoustical conditions may violate the basic assumptions of SDM of plane waves, thus one should not expect that the reproduced sound field is an exact replica of the recorded. The current SDM assumes single reflection per analysis window, which may not be always the case in a car cabin. Nevertheless, both objective and perceptual results suggest that SDM preserves the perceptual differences between the stimuli set in the experiment - a vital requirement for perceptual assessment. The advantages of the method rely on the flexible reproduction scheme, the fast acquisition of VIR compared to BCS, as well as analysis and visualization capabilities of the spatial properties of the VIRs.

Further work on this topic aims to understand the perceptual effects of acoustical properties (i.e. reverberation) in everyday listening spaces, such domestic rooms and car-cabins. The above methodology is currently followed in assessing car-cabins in detail, aiming to identify the perceptual constructs originating from acoustical alterations in the sound field in question. Moreover a database of multiple car cabins has been established to be perceptually evaluated, aiming to provide insights in the perceptual characterization between different-sized or type of car cabins, as well as sound systems within similar cabins, and the effect of human occupancy [4]. The proposed methodology will also be applied in small room acoustics aiming to perceptually assess a variety of acoustical settings in standard listening rooms and to better understand the interaction between the acoustical properties of the reproduction room that are inherently imposed on the reproduced sound field.

The experimental methodology proposed here, is an example of an application in car audio evaluation. However, the method could be applied in many domains. Since SDM is a parametric approach, certain applications could be realized by the using the suggested system, such as development of tuning tool for a car audio system in the lab, as well as sound and system design testing algorithms, without the need of a prototype car.

Acknowledgments:

The authors would like to thank their colleagues at Bang & Olufsen, Aalborg University and Aalto University for their support and helpful input. The research leading to these results has received funding from the European Union's Seventh Framework Program (FP7/2007-2013) under grant agreement no. ITN-GA-2012-316969.

References

[1] P. Hegarty, S. Choisel, and S. Bech. A Listening Test System for Automotive Audio - Part 3: Comparison of Attribute Ratings Made in A Vehicle with those made using an Auralization System. In *123rd Convention of the Audio Engineering Society*, New York, USA, 2007.

[2] G. Martin and Søren Bech. Attribute Identification and Quantification in Automotive Audio - Part 1: Introduction to the Descriptive Analysis Technique. In *118th Convention of the Audio Engineering Society*, Barcelona, Spain, 2005.

[3] S. Bech, M.-A. Gulbol, G. Martin, J. Ghani, and W. Ellermeier. A listening test system for automotive audio – Part 2- Initial verification. In *118th Convention of the Audio Engineering Society*, Barcelona, Spain, 2005.

[4] F. Christensen, G. Martin, M. Lydolf, M. Pauli, S. Woo-Keun, P. Benjamin, and M. Lydolf. A Listening Test System for Automotive Audio - Part 1 : System Description. In *118th Convention of the Audio Engineering Society*, Barcelona, Spain, 2005.

[5] S. E. Olive. A New Reference Listening Room For Consumer, Professional and Automotive Audio Research. In *126th Convention of the Audio Engineering Society*, Munich, Germany, 2009.

[6] R. Shively. Subjective Evaluation of Reproduce d Sound in Automotive Spaces. In *15th International Conference of Audio Engineering Society*, Copenhagen, Denmark, 1998.

[7] N. Otto. Listening Test Methods for Automotive Sound Quality. In *103rd Convention of the Audio Engineering Society*, New York, USA, 1997.

[8] A. Azzali, A. Farina, G. Rovai, G. Boreanaz, and G. Irato. Construction of a Car Stereo Audio Quality. In *117th Convention of the Audio Engineering Society*, San Fransisco, USA, 2004.

[9] E. Ugolotti, G. Gobbi, and A. Farina. IPA - A Subjective Assessment Method of Sound Quality of Car Sound Systems. In *110th Convention of the Audio Engineering Society*, Amsterdam, The Netherlands, 2001.

[10] I. S. Organisation. ISO 3382-2:2008, Acoustics—Measurement of Room Acoustic Parameters—Part 2: Reverberation Time in ordinary Rooms (International Standards Organization, Geneva, Switzerland, 2009). 2008.

[11] I. S. Organization. ISO: 3382-1:2009, Acoustics—Measurement of Room Acoustic Parameters—Part 1: Performance Spaces (International Standards Organization, Geneva, Switzerland, 2009). 3, 2009.

[12] N. Kaplanis, S. Bech, S. H. Jensen, and T. V. Waterschoot. Perception of Reverberation in Small Rooms : A Literature Study. In *55th International Conference of the Audio Engineering Society on Spatial Audio*, Helsinki, Finland, 2014.

[13] ITU-R. Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems: BS.1116-1. *International Telecomm Union, Geneva, Switzerland*, pages 1–26, 1997.

[14] International Telecommunication Union. P.800: Methods for subjective determination of transmission quality. *ITU-T Recommendation*, 800, 1996.

[15] J. Delarue, B. J. Lawlor, and M. Rogeaux. *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research.* 2014.

[16] S. E. Olive and T. Welti. Validation of a Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems. *36th International Conference: Automotive Audio*, pages 1–7, 2009.

[17] W. L. Martens, H. Sakanashi, and W. Woszczyk. Whole-body vibration associated with low-frequency audio reproduction influences preferred equalization. In *36th International Conference of the Audio Engineering Society*, Dearborn, USA, 2009.

[18] S. Germain, S. E. Olive, and T. Welti. The Effect of Whole-body Vibrations on Preferred Bass Equalization of Automotive Audio Systems. In *127th Convention of the Audio Engineering Society*, New York, USA, 2009.

[19] M. Sams, R. Hari, J. Rif, and J. Knuutila. The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence. *Journal of Cognitive Neuroscience*, 5(3):363–368, Dec. 1993.

[20] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki. Spatial decomposition method for room impulse responses. *Journal of the Audio Engineering Society*, 61(1/2):17–28, 2013.

[21] T. Lokki, J. Patynen, A. Kuusinen, and S. Tervo. Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *Journal of the Acoustical Society of America*, 132(5):3148–3161, 2012.

[22] S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki. Preferences of critical listening environments among sound engineers. *AES: Journal of the Audio Engineering Society*, 62(5):300–314, 2014.

[23] S. Parkin, G. M. Mackay, and A. Cooper. How Drivers Sit in Cars. *Accident Analysis & Prevention*, 27(6):777–783, 1995.

[24] A. Farina. Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique. In *108th Convention of The Audio Engineering Society*. Audio Engineering Society, Feb. 2000.

[25] S. Tervo, J. Pätynen, N. Kaplanis, S. Lydolf Morten, Bech, and T. Lokki. Spatial Analysis and Synthesis of Car Audio System and Car Cabin Acoustics with a Compact Microphone Array. *J. Audio Eng. Soc.*, 63(11):914–925, 2015.

[26] V. Pulkki. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society*, 45(6):456–466, June 1997.

[27] J. Patynen, S. Tervo, and T. Lokki. Amplitude panning decreases spectral brightness with concert hall auralizations. In *55th International Conference of the Audio Engineering Society on Spatial Audio*, Helsinki, Finland, 2014.

[28] J. Pätynen, S. Tervo, and T. Lokki. Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses. *The Journal of the Acoustical Society of America*, 133(2):842–57, 2013.

[29] S. Bech and N. Zacharov. *Perceptual audio evaluation-Theory, method and application*. Wiley, 2007.

[30] F. Rumsey. Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society*, pages 651–666, 2002.

[31] F. E. Toole. *Sound Reproduction: Loudspeakers and Rooms.* Focal Press, Bostom, 2008.

[32] E. Granier, M. Kleiner, B. Dalenbäck, and P. Svensson. Experimental Auralization of Car Audio Installations. *Journal of the Audio Engineering Society*, 44(10):835–849, 1996.

[33] International Organization for Standardization. ISO 3745::2003 - Acoustics. Determination of sound power levels and sound energy levels of noise sources using sound pressure. Precision methods for anechoic rooms and hemi-anechoic rooms. *International Standard Organisation*, 2003.

[34] H. T. Lawless and H. Heymann. *Sensory Evaluation of Food: Principles and Practices.* 1999.

[35] T. Lokki, J. Pätynen, and N. Zacharov. Concert Hall Acoustics Assessment with Sensory Evaluation - Tools and Practices. *Prooceedings of the Institute of Acoustics*, 33(2):35–39, 2011.

[36] A. Kuusinen, J. Paätynen, S. Tervo, and T. Lokki. Relationships between preference ratings, sensory profiles, and acoustical measurements in concert halls. *The Journal of the Acoustical Society of America*, 135(1):239–250, Jan. 2014.

[37] S. Choisel and F. Wickelmaier. Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. *The Journal of the Acoustical Society of America*, 121(1):388, 2007.

[38] J. Francombe, R. Mason, M. Dewhirst, and S. Bech. Elicitation of attributes for the evaluation of audio-on-audio interference. *The Journal of the Acoustical Society of America*, 136(5):2630–2641, 2014.

[39] J. Berg and F. Rumsey. Spatial attribute identification and scaling by repertory grid technique and other methods. *16th International Conference: Spatial Audio*, 1999.

[40] J. Berg and F. Rumsey. Identification of quality attributes of spatial audio by repertory grid technique. *Journal of the Audio Engineering Society*, pages 365–379, 2006.

[41] T. Neher, T. Brookes, and F. Rumsey. A Hybrid Technique for Validating Unidimensionality of Perceived Variation in a Spatial Auditory Stimulus Set. *Journal of the Audio Engineering Society*, 54(4):259–275, Apr. 2006.

[42] G. Lorho. Perceived Quality Evaluation - An application to sound reproduction over headphones. *Phd Thesis*, 2010.

[43] C. Simonsen and S. Legarth. A procedure for sound quality evaluation of hearing aids. *Hearing Review, Dec*, 17(13):32–37, 2010.

[44] N. Zacharov, J. Ramsgaard, G. Le Ray, and C. V. Jorgensen. The multidimensional characterization of active noise cancellation headphone perception. In *Second International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 130–135. IEEE, June 2010.

[45] D. Valentin, S. Chollet, M. Lelièvre, and H. Abdi. Quick and dirty but still pretty good: A review of new descriptive methods in food science. *International Journal of Food Science and Technology*, 47(8):1563–1578, 2012.

[46] J. Delarue and J. M. Sieffermann. Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*, 15(4):383–392, 2004.

[47] ITU-R. Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA) International. *International Telecommunication Union Recommendations BS. 1534-1*, 2001.

SUMMARY

The central topic of this thesis is Reverberation. The reported studies focus on the effects of reverberation that are likely to occur in common listening environments, such as car cabins and ordinary residential listening rooms.

The experimental designs followed in this work made use of a novel assessment framework, which forms a significant part of this thesis. The proposed framework overcomes previously identified challenges in perceptual evaluation of room acoustics, relating to acquisition and presentation of the acoustical fields, as well as the perceptual evaluation of such complex sound stimuli.

Overall the work described in this thesis contributes to: (1) understanding the perceptual effects imposed in the reproduced sound within automotive and residential enclosures, and (2) the design and implementation of a perceptual assessment protocol for evaluating room acoustics.

ISSN (online): 2446-1628
ISBN (online): 978-87-7112-869-7

AALBORG **UNIVERSITY PRESS**