Timbre Models of Musical Sound

From the model of one sound to the model of one instrument

Jensen, Karl Kristoffer

Publication date:
1999

Document Version
Også kaldet Forlagets PDF

Rapport
Nr. 99/7
ISSN 0107-8283

Timbre Models of Musical Sounds

Kristoffer Jensen

Datalogisk Institut, Københavns Universitet
Department of Computer Science, University of Copenhagen
Universitetsparken 1, DK-2100 København Ø
*** DIKU * TRYK ***

Timbre Models of Musical Sounds

From the model of one sound

to the model of one instrument

Kristoffer Jensen

Ph.D. dissertation under the supervision of **Jens Arnspang**

In memory of

H. E. Eddy

Erik H. Nielsen

CR Subject Classification : H-5-5

Abstract

This work involves the analysis of musical instrument sounds, the creation of timbre models, the estimation of the parameters of the timbre models and the analysis of the timbre model parameters.

The timbre models are found by studying the literature of auditory perception, and by studying the gestures of music performance.

Some of the important results from this work are an improved fundamental frequency estimator, a new envelope analysis method, and simple intuitive models for the sound of musical instruments. Furthermore a model for the spectral envelope is introduced in this work. A new function, the brightness creation function, is introduced in the spectral envelope model.

The timbre model is used to analyze the evolution of the different timbre parameters when the fundamental frequency is changed, but also for different intensity, tempo, or style. The main results from this analysis are that brightness rises with frequency, but nevertheless the fundamental has almost all amplitude for the high notes. The attack and release times generally fall with frequency. It was found that only brightness and amplitude are affected by a change in intensity, and only the sustain and release times are affected when the tempo is changed.

The different timbre models are also used for the classification of the sounds in musical instrument classes with very good results. Finally, listening tests have been performed, which assessed that the best timbre model has an acceptable sound quality.

Resumé

Dette arbejder omhandler analyse af musikinstrumenter, dannelse af modeller af musikinstrumenters klangfarve, estimering af klangfarve model parametre og analyse af modelparametrene.

Klangfarvemodellerne er fundet ved at gennemgå lydperceptorisk litteratur, og ved at studere musikudøvelse.

Nogle vigtige resultater fra dette arbejde er en forbedret fundamental frekvens estimator, en ny envelope analysemetode, og simple intuitive modeller af musiklyd. Desuden er en model af den spektrale envelope udviklet. I den forbindelse er en ny funktion for syntese af lyd med en given ‘brightness’ udviklet.

Klangfarvemodellen er brugt til at analysere udviklingen af de forskellige klangfarveattributter, når fundamentalfrekvensen ændres, men også for forskellige

intensiteter, tempi og stil. De vigtigste konklusioner fra dette arbejde er, at ‘brightness’ stiger med frekvens; men fundamentalen har alligevel næsten al amplitude for de høje toner. ‘Attack’ og ‘release’ tiderne falder med frekvensen. Af intensitets- og tempoændringer fandtes, at kun ‘brightness’ og amplituden ændres når intensiteten ændres, og at kun ‘sustain’ og ‘release’ tiderne ændres når tempoet ændres.

De forskellige klangfarvemodeller er også brugt til klassifikation af lyd i instrumentklasser med meget godt resultat. Lytteforsøg godtgjorde, at den bedste klangfarvemodel har en acceptabel lydkvalitet.

Résumé

Ce travail traite l’analyse des sons musicaux, la création des modèles de timbre, l’estimation des paramètres des modèles de timbre, ainsi que l’analyse des paramètres des modèles.

Les modèles de timbre ont été trouvés dans la littérature de la perception auditive et en étudiant les gestes du musicien.

Quelques résultats importants du travail présenté ici sont une estimation améliorée de la fréquence fondamentale. Une nouvelle méthode pour l’estimation des temps d’attaque et de relâchement a été developpée, ainsi que des modèles intuitifs de sons d’instrument de musique. Un nouveau modèle d’enveloppe spectrale a été défini, ainsi qu’une fonction qui donne un son avec la brillance indiquée.

Les modèles de timbre sont utilisés pour l’analyse de l’évolution des paramètres des timbres en fonction de la fréquence fondamentale, de l’intensité, du tempo ou du style. Le résultat principal de cette analyse est que la brillance monte avec la fréquence, mais que la fondamentale a presque toute l’amplitude dans les aigüs. Les temps d’attaque et relâchement diminuent avec la fréquence fondamentale. Pour une variation de l’intensité, seul l’amplitude et la brillance sont affectées. Seuls les temps de maintien et relâchement changent avec le tempo.

Le modèle de timbre est aussi utilisé pour la classification des sons dans des classes d’instruments avec de très bons resultats. Finalement, des tests d’ecoute de tous les modèles ont permis de conclure que le meilleur modèle de timbre possède une qualité de son acceptable.

Acknowledgments

First and foremost, my thanks go to Jens Arnspang, who has created the music informatics group at the Computer Science Department at the University of Copenhagen, and without whom this work would never have started. Jens accepted to be my supervisor and had the open mind to let me pursue my own directions, and detours.

Secondly, my thanks go to the two members of the monitor group, Ivar Frounberg and Holger Rindel for insightful comments and feedback both in the musical and the technical domain. The comments from them helped keep a focus in my work, and inspired further improvements.

This work has been financed by the Danish Technical Research Council whom I thank.

My thanks go to all members, past or present, of the music informatics group. Special thanks go to Klaus Hansen for invaluable help. Fruitful discussions with Stefan Borum, Anders Møller and Esben Skovenberg have also been a great source of inspiration.

Sincere thanks goes to the musicians who accepted to spend time to record sounds *which is not music*. The musical instrument sound database created with their help has been instrumental in this work.

The judgments and comments from the members of the listening tests have also been a great help. My sincere thanks go to all the participants in the listening tests.

Many helpful comments have also come from other groups in the computer science department. Special thanks to Ketil Perstrup, Kristian Pilgaard, Jon Sporring, Joachim Weickert, Peter Riber, Stig Skelboe, Knud Henriksen, Erik Frøkjær and Morten Hanehøj. The image group and notably the scale-space community have been a great source of inspiration.

My thoughts go to everybody at DIKU who have made my stay here so pleasant.

Part of the thesis work in Denmark is passed in a different research institution, as required by the Danish Ph.D. circular. I was very lucky to be accepted at the Groupe Informatique Musical at the Laboratoire Mecanique et Acoustique in Marseille, France. My sincere thanks go to Jean-Claude Risset for having accepted me in his group, and to Richard Kronland-Martinet and Philippe Guillemain for help and discussions. My stay at the groupe informatique musicale was made agreeable by the fruitful discussions and the nice atmosphere in the group.

A final thanks goes to Carol Jensen, Thomas Jensen and all the members of my family, and especially to Alice, who hopefully will see more of Papa soon.

Table of Contents

1. INTRODUCTION 1

1.1. FRAMEWORK 2
1.2. WORK METHODOLOGY 4
1.3. STRUCTURE OF THE DOCUMENT 5

2. MUSICAL INSTRUMENTS 7

2.1. INTRODUCTION 7
2.2. CONTROL 8
2.3. TIMBRE DIMENSIONS 10
2.3.1. Identity 10
2.3.2. Pitch, Loudness and Duration 11
2.3.3. Dissimilarity Tests 12
2.3.4. Verbal Attributes 13
2.3.5. Noise 13
2.3.6. Roughness 13
2.4. ADDITIVE MODEL 14
2.4.1. Time-Frequency Analysis 15
2.4.2. Phase 15
2.5. DATABASE 16
2.6. CONCLUSIONS 17

3. FUNDAMENTAL FREQUENCY ESTIMATION 19

3.1. INTRODUCTION 19
3.2. FFT CANDIDATES 20
3.2.1. Frequency and Amplitude Estimation 21
3.2.2. Masking 22
3.3. FUNDAMENTAL FREQUENCY ESTIMATION 23
3.3.1. Frequency Difference Fundamental Estimation 24
3.3.2. Missing Frequencies 25
3.3.3. Fit Stretched Harmonic Curve. 26
3.4. INITIAL FREQUENCIES 27
3.4.1. Harmonic Frequencies 27
3.4.2. Spurious Partials 27
3.4.3. Spectrogram Analysis 28
3.5. PITCH TRACKER 28
3.5.1. Moving Fundamental Frequency 29
3.5.2. Instantaneous Frequency 29

3.5.3. Curve Segmentation *30*
3.6. CONCLUSIONS 31

4. ANALYSIS/SYNTHESIS 33

4.1. INTRODUCTION 34
4.2. FAST FOURIER TRANSFORM BASED ADDITIVE ANALYSIS 35
4.2.1. Sliding Window Analysis *35*
4.2.2. Better Timing Resolution *35*
4.2.3. Partial Track *36*
4.2.4. FFT Conclusions. *37*
4.3. LINEAR TIME/FREQUENCY ANALYSIS 38
4.3.1. Constructing the Filters *39*
4.3.2. Initial Frequencies *42*
4.3.3. Rebounds *42*
4.3.4. Frequency and Amplitude Extraction *44*
4.3.5. Data Reduction *45*
4.4. COMPARISON OF FFT AND LTF ANALYSIS 45
4.4.1. Test Signals *45*
4.4.2. Analysis *45*
4.4.3. Results *46*
4.5. RESYNTHESIS 47
4.6. CONCLUSIONS 48

5. ENVELOPE MODELING 51

5.1. INTRODUCTION 52
5.2. TIMING EXTRACTION 54
5.2.1. Percent Method *54*
5.2.2. Slope Method *56*
5.2.3. Percent vs. Slope *59*
5.2.4. Relative Amplitude (percents) *59*
5.3. CURVE FORM 60
5.3.1. Curve Model *60*
5.3.2. Language Conventions *61*
5.3.3. Curve Fitting *62*
5.4. RECONSTRUCTION OF THE ENVELOPE 63
5.5. RECREATION OF THE ADDITIVE PARAMETERS 64
5.6. ENVELOPE SHARPENING 65
5.7. CONCLUSION 66

6. HIGH LEVEL ATTRIBUTES 67

6.1. INTRODUCTION 68
6.2. ADDITIVE PARAMETER ANALYSIS 69
6.3. SPECTRAL ENVELOPE 69
6.4. FREQUENCY 70
6.5. ENVELOPE 71
6.5.1. Timing Analysis *71*
6.5.2. Curve Form Analysis *73*
6.6. NOISE 74
6.6.1. Distribution of Partial Noise *74*
6.6.2. Spectrum of Partial Noise *75*
6.6.3. Correlation of Partial Noise *76*
6.6.4. Resynthesis of Noise *78*
6.6.5. Noise Conclusion *78*
6.7. HLA VISUALIZATION 79
6.8. RECREATION OF THE ADDITIVE PARAMETERS 81
6.9. CONCLUSION 84

7. SPECTRAL ENVELOPE MODEL 85

7.1. INTRODUCTION 86
7.2. ANALYSIS OF PERCEPTIVE ATTRIBUTES 87

7.2.1. Brightness ... *88*
7.2.2. Time domain Brightness Function ... *89*
7.2.3. Tristimulus ... *92*
7.2.4. Odd/Even Relation ... *93*
7.2.5. Irregularity ... *93*
7.3. SPECTRAL ENVELOPE MODEL ... 94
7.3.1. The High Harmonic Components ... *95*
7.3.2. The Low Harmonic Components ... *95*
7.3.3. Finding the Positive Range ... *96*
7.3.4. Finding Best Irregularity ... *97*
7.3.5. Recreation of Spectral Envelope ... *98*
7.4. TIME VARYING SPECTRAL ENVELOPE ... 99
7.5. FORMANTS ... 101
7.6. CONCLUSION ... 103

8. MINIMAL DESCRIPTION ATTRIBUTES ... 105

8.1. INTRODUCTION ... 105
8.2. FREQUENCY MODEL ... 107
8.3. AMPLITUDE MODEL ... 108
8.4. GENERIC PARAMETER MODEL ... 109
8.4.1. Envelope Parameters ... *109*
8.4.2. Noise Parameters ... *113*
8.4.3. Comments on the Noise Model ... *114*
8.5. ERROR TERM CALCULATION ... 115
8.6. ANALYSIS FROM HLA ATTRIBUTES ... 117
8.7. RECREATION OF HLA ATTRIBUTES ... 117
8.8. SOUND SYNTHESIS FROM THE MDA ... 120
8.9. CONCLUSION ... 122

9. INSTRUMENT DEFINITION ATTRIBUTES ... 123

9.1. INTRODUCTION ... 124
9.2. HALF OCTAVE BANDS ... 125
9.3. IDA PARAMETER CALCULATION ... 127
9.4. IDA CLASSES ... 128
9.5. FUNDAMENTAL FREQUENCY EVOLUTION ... 128
9.5.1. Spectral Envelope Evolution ... *128*
9.5.2. Frequency Parameter Evolution ... *132*
9.5.3. Envelope Evolution ... *133*
9.5.4. Noise Evolution ... *135*
9.6. LOUDNESS ... 138
9.6.1. Spectral Envelope Parameters ... *138*
9.6.2. Frequency Parameters ... *139*
9.6.3. Envelope Parameters ... *140*
9.6.4. Noise Parameters ... *141*
9.6.5. Loudnesses conclusions ... *143*
9.7. TEMPO ... 143
9.7.1. Spectral Envelope Parameters ... *143*
9.7.2. Frequency Parameters ... *144*
9.7.3. Envelope Parameters ... *145*
9.7.4. Noise Parameters ... *145*
9.7.5. Tempo Conclusions ... *146*
9.8. STYLE ... 146
9.8.1. Spectral Envelope Parameters ... *147*
9.8.2. Frequency Parameters ... *148*
9.8.3. Envelope Parameters ... *148*
9.8.4. Noise Parameters ... *150*
9.8.5. Style Conclusions ... *150*
9.9. SOUND RECREATION FROM IDA PARAMETERS ... 151
9.10. CONCLUSIONS ... 151

10. TIMBRE MODIFICATIONS ... 153

10.1. INTRODUCTION 153
10.2. PITCH, LOUDNESS AND DURATION 155
10.2.1. Pitch *155*
10.2.2. Loudness *156*
10.2.3. Duration *156*
10.2.4. Number of Partials *157*
10.3. INTER-MODEL MODIFICATIONS 158
10.4. CONCATENATION 159
10.4.1. Superposition *160*
10.4.2. Replacement *160*
10.5. ADDITIVE MODIFICATIONS 161
10.5.1. Spectral Envelope *162*
10.5.2. Frequency *162*
10.5.3. Envelope *163*
10.5.4. Noise Modification *166*
10.5.5. Verification *170*
10.6. RESYNTHESIS 171
10.7. CONCLUSIONS 171

11. VERIFICATION OF THE TIMBRE MODELS 173

11.1. INTRODUCTION 174
11.2. SOUNDS 175
11.3. TIMBRE ATTRIBUTES 175
11.3.1. HLA *175*
11.3.2. MDA *176*
11.3.3. IDA *177*
11.4. NYQUIST FREQUENCY AMPLITUDE 177
11.5. PRINCIPAL COMPONENT ANALYSIS 179
11.6. CLASSIFICATION 182
11.7. CONCLUSIONS 183

12. LISTENING TESTS 185

12.1. INTRODUCTION 185
12.2. RATING SCALES 186
12.3. ORIGINAL SOUNDS 187
12.4. MODEL SOUNDS 187
12.5. LISTENING PANEL 188
12.6. TRAINING 188
12.7. TEST PROCEDURE 188
12.8. SUBJECT COMMENTS 189
12.8.1. The Test Procedure *189*
12.8.2. The Impairment Scale *189*
12.8.3. The Sounds *190*
12.9. STATISTICAL PRESENTATION 190
12.9.1. Model Degradation *191*
12.9.2. Instrument Degradation *191*
12.9.3. Subject Scores *192*
12.9.4. Analysis/Synthesis Instrument Degradation *192*
12.9.5. Degradation as a Function of Fundamental Frequency *193*
12.9.6. The HLA Instrument Degradations *194*
12.9.7. MDA Instrument Degradation *194*
12.9.8. The IDA Instrument Degradations *195*
12.9.9. Model Degradation with Soprano Removed *195*
12.9.10. Complete Scores *196*
12.10. CONCLUSIONS 197

13. CONCLUSIONS 199

13.1. THE TIMBRE MODELS 199
13.2. TIMBRE MODIFICATIONS 202
13.3. TIMBRE MODEL EVALUATION 203
13.4. FUTURE DIRECTIONS 204

13.4.1. Parameter Estimation *204*
13.4.2. Model Scope *205*

14. REFERENCES **207**

15. TABLE OF FIGURES **219**

A. SOUND RECORDINGS **A-1**

A.1. VIOLIN A-1
A.2. VIOLA A-2
A.3. CELLO A-3
A.4. SAXOPHONE A-4
A.5. CLARINET A-5
A.6. FLUTE A-6
A.7. SOPRANO A-7
A.8. PIANO A-8

B. LISTENING TEST INSTRUCTIONS IN DANISH **B-1**

Chapter 1

1. Introduction

The initial inspiration for this work was the need to understand the transitions of musical sounds. The transition was soon defined as being the variation over time of pitch, loudness and timbre, and the classification of these variations. [Strawn 1985] offers further insight on the transitions of musical instruments. Pitch and loudness are fairly well known parameters, but timbre is less well defined, although generally defined as multi-dimensional.

Timbre then naturally became the main subject of this work. Two approaches were tested to understand the dimensions of timbre, the first by examining the physical gestures associated with playing an instrument and the other by looking at the perception and psychoacoustic literature. This can be seen as a global approach, encompassing both the performer of a musical instrument and the auditor of the sounds produced. The conclusions of the two approaches were then used in the analysis and modeling of musical instrument sounds.

The analysis of transitions was eventually left out, and the work is now done on isolated musical instrument sounds. The goal is to find a few parameters which are relevant to human perception and which model music sounds well. Furthermore, the evolution of

sounds, as a function of playing style, loudness, or note played, should also be well modeled. Ideally, this would equal a musical instrument, but much work remains before this goal is achieved. Instead, this work is the basis for a better understanding of what timbre is, and also the basis for a digital musical instrument with potentially the same timbre quality and versatility as an acoustic instrument, in expression as good as the best acoustical instruments.

The model of musical sounds presented here can be used as a basis for compression of (musical) sounds, for interactive distributed music, or for research in composition with timbre. For a survey on timbre composition, see for instance [Barrière *et al.* 1991].

In general terms, musical informatics research can be helpful for classical music research, for auditory perception research and for the auditory display research. Fundamental methods developed in the music informatics community can potentially find uses in any domain.

1.1. Framework

This work balances on the border between analysis and synthesis of sounds of musical instruments, which can be seen as an example of analysis by synthesis [Risset 1991].

Analysis is done on sounds, but also on the parameters of preceding analysis. This is done so that the important timbre attributes of a sound will emerge. The last model will present some parameters which are important timbre attributes, but which in an automatic framework, can not (yet) resynthesize an acceptable sound. However, this is believed to be more a problem with the estimation of the parameters of the models than with the models themselves. Therefore, it is believed that the models can be used to synthesize good quality sounds, if the parameters are adjusted appropriately.

Each model has an inverse function, which allows one to recreate the input parameters from the output parameters. The recreation is never identical, and some of the perceptual loss can be found by studying the listening test results in Chapter 12.

The different steps of the analysis/model/synthesis can be seen in figure 1.1. The sounds are first analyzed into additive parameters, where sinusoidals, called partials, with time-varying amplitude and frequency are added together. The sinusoidals correspond often, although not always, to the fundamental and the harmonic overtones of the sound being analyzed. Then the partials are analyzed, and a few perceptually important parameters are found and stored in the High Level Attribute (HLA) model. This is done for each partial.

In the Minimum Description Attribute (MDA) model, the parameters of the HLA model are defined by the fundamental value and the evolution over partial index. Finally, the Instrument Definition Attribute (IDA) model includes the MDA parameters for the full playing range of an instrument. The IDA model is therefore a collection of many MDA sets.

In the MDA and the IDA models, the partials need to be quasi-harmonic. This is not the case for the additive and the HLA models.

All models have an inverse function, which permits recreating the previous level parameters all the way to the resynthesis of the sound.

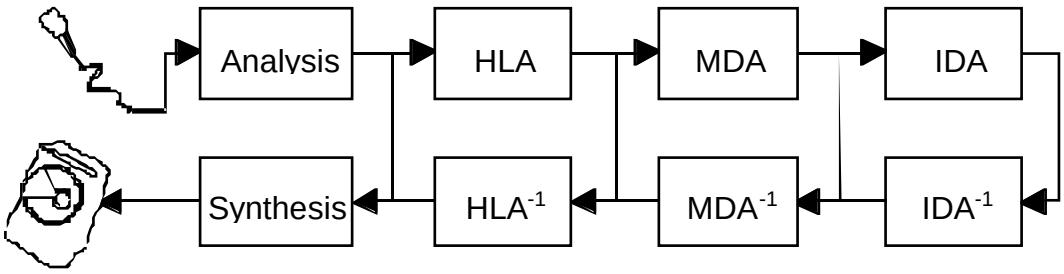


Figure 1.1. Complete flow chart of analysis and modeling in this work.

Visualization of the additive parameters is useful when a view of the general shape of the sound is needed. The HLA parameters are useful when the timbre attributes, such as the attack time or the brightness of a sound, need to be visualized. The MDA model introduces a model of the spectral envelope. The MDA model is assumed to contain all the information of a sound in the fewest possible parameters.

The IDA model parameters are useful when the difference between instruments, or between expressions of the same instrument, needs to be analyzed or visualized.

Furthermore, the validity of each model can be estimated by the ability of the parameters of the model to classify the sounds in instrument families. Some experiments on the classification have been performed in the validation of the timbre models presented in Chapter 11 with good results.

1.2. Work Methodology

The first part of this work consisted in finding expressions of musical instruments. This work was conducted by interviewing musicians, and recording musical instruments in as many expressions as possible.

When the goal of this work was restated into finding a model for the timbre of musical instruments, an iterative process of finding the parameters of such a model began. The parameters of the model are of course very dependent on the analysis model of the sound. The analysis model was therefore first defined to be additive.

The additive parameters generally model only the voiced part of the sound, and the noise analysis should therefore be found. The use of a better additive analysis method allows the choice of the less frequently used model of noise using the irregularity of the additive parameters.

When the analysis parameters were chosen, the analysis of musical instrument sounds could begin. Quality of the analysis was judged by listening to the resynthesized sounds and by analyzing the resulting additive parameters. At the same time, the timbre model was initiated. This was done by experimenting with simple models of the additive parameters, and by studying the auditory perception literature. The quality of the timbre models was evaluated by listening to the resynthesis of the sounds from the models, and by analyzing the parameters of the model. The initial analysis and the first timbre model, the HLA model, were changed if necessary. Furthermore, new musical sounds were recorded, if another dimension of the timbre space was to be evaluated. Then the simpler timbre model, the MDA model, was initiated and the process was repeated, now including another level.

Finally, the full instrument model, the IDA model, was introduced. Now the parameters could be analyzed as a function of the playing range, or other expressive scales. The underlying models were evaluated on the basis of this analysis, and changed when necessary. Furthermore, listening tests were performed, and classification experiments using the timbre models were also performed. All this gave rise to more modification of the timbre models, after which the quality of each model was again evaluated.

This ascending methodology was necessary, since no timbre models were found in the literature. The deductive conclusions are not strictly speaking unique. Nevertheless, this

methodology is believed to be the best for this work. The relatively dispersive literature search has facilitated finding better models and better foundations for the models chosen.

Conclusive timbre models with promising applications are introduced in this work.

1.3. Structure of the Document

Chapter 2 presents the musical instruments, the control and perception of musical sounds, the timbre and the additive model. Chapter 3 introduces an improved fundamental frequency estimator, and the estimation of the initial frequencies used in the analysis chapter. Chapter 4 explains the analysis of the additive parameters and compare two methods, the well-known FFT-based analysis, and a new analysis method, developed by Philippe Guillemain [Guillemain *et al.* 1996], based on a linear sum of gaussian kernels. The conclusion is that the new analysis method, here called the LTF analysis, has a time resolution that is twice as good as the optimal two-pass FFT-based analysis.

Chapter 5 explains the envelope model and compares two methods for the extraction of envelope times: the first, which finds the envelope times at a certain percentage of the maximum amplitude, and a new method developed here, which finds the envelope times by analyzing the derivative of the amplitude envelope. This method, which is called the slope method, performs significantly better than the simpler percent-based method. Chapter 6 introduces the HLA model, which models the sound with a few perceptually relevant parameters for each partial: spectral envelope, mean frequencies, envelope, and amplitude and frequency irregularities (shimmer and jitter).

Chapter 7 introduces the spectral envelope model used in the MDA model that is presented in Chapter 8. The spectral envelope model parameters include brightness, and a function for the creation of a signal with a given brightness is given in the additive and in the time domain. The MDA model is based on the HLA model, but it further models the partial evolution for each parameter.

Chapter 9 introduces the IDA model, which is a model for the evolution of the MDA parameters as a function of the fundamental frequency. This chapter also discusses the evolution of the timbre attributes as a function of fundamental frequency, intensity, tempo or style. Several important results of this analysis are given in Chapter 9.

Chapter 10 introduces the timbre modifications of the different timbre models. Chapter 11 examines the validity of the timbre models by classification methods. The result is that the timbre attributes can classify 150 sounds from the full playing range of five

instruments with no errors. Chapter 12 verifies the validity of the resynthesis of the timbre models by performing listening tests. Chapter 13, finally, offers a conclusion and a proposal for further work.

Chapter Two

2. Musical Instruments

In this chapter the musical instrument is presented from the two most common points of view, the gestural, and the perceptive. The gestural point of view discusses the playing of an instrument, while the perceptive point of view discusses the perception involved in listening to musical instrument sounds. Based on some initial research into the control of musical instruments, a database of musical instrument sounds has been created. Furthermore, the model of the sound of the musical instrument is presented here. The conclusion of the perceptive research reviews is the basis of the timbre models in the following chapters.

2.1. Introduction

A model of musical instruments should obey two fundamental obligations. It needs good sound quality and easy control of the important expression attributes.

This chapter investigates the literature on auditory perception, timbre analysis and control of musical instruments. The conclusions from this chapter are used in the following

chapters to create the models of musical instruments. The discussions of musical instruments have also been important for the choice of musical instruments that are used in the analysis of the timbre models. The control of musical instruments is investigated by analyzing the current situation and proposals for future systems of digital musical instrument interfaces. Some results from the research on reaction time from different stimuli are also given.

The timbre conclusions are given from a review of auditory perception literature and from verbal attribute research.

The musical instruments being analyzed in this work are the quasi-harmonic instruments. The term quasi-harmonic denotes instruments whose partial frequencies are close to harmonic. This means that for example the drums, cymbals, and carillons have been excluded.

The actual instruments being analyzed have been chosen for the quality of expression, for general recognition, and for availability.

In this chapter the control of musical instruments is discussed in section 2.2, then the timbre of musical sounds is discussed in section 2.3. The additive model of musical sounds is presented in section 2.4, with a discussion of the phase sensitivity in paragraph 2.4.2. The database of musical instrument is discussed in section 2.5. Finally a conclusion is offered.

2.2. Control

The control of a musical instrument is here defined to be the physical process of moving or manipulating the parts of the musical instrument to produce sounds. The analysis of the control of musical instruments was done in an early stage of this work and only summarized here. Some general reflections on the control of musical instruments can be found in [Jensen 1996a], and an overview of the control of the violin can be found in [Jensen 1996b]. This research is the basis for the constitution of the database of musical instrument sounds, and the classification of the sounds in families of intensity, style, or other parameters, such as the speed of the bow of the violin.

In mainstream computer-based music, control is generally achieved with the Musical Instruments Digital Interface (MIDI) interface [IMA 1983]; most often through a piano like Midi Master Keyboard [Jensen 1988].

[Moore 1988] criticized the “degree of control intimacy” of MIDI. Several replacements have been proposed without success, see for instance [ZIPI 1994].

Much other work in control of musical instruments, or gesture research, has been done. [Vertegaal *et al.* 1996] stresses the importance of a “tight relationship between the musician and the instrument.” [Wanderley *et al.* 1998] present their work in gestural research, as well as the gestural research discussion group, which they manage.

A system which is perhaps comparable to acoustic instruments is presented in [Cadoz *et al.* 1984], [Cadoz *et al.* 1990]. The haptic interface, which gives sensory feedback to the performer, seems to enhance intimacy considerably.

[Jensen 1996a] argues that even though there are many dimensions to the control of a musical instrument, the performer concentrates only on a few of the controls at any given time. An argument for or against this hypothesis can perhaps be found in the literature on human reaction time. [Leonard 1959] did a much-cited work in which he studied the reaction time when one or several fingers were stimulated with a 50 Hz vibration. His results show “a difference between simple reaction time and two-choice times, but no systematic differences between 2, 4, or 8 choices.” This would imply that a human could react to 8 choice stimuli just as fast as to 2 choice stimuli. His results were not replicated in a later study, [Hoopen *et al.* 1981], which shows that the reaction time increases with the number of choices. This increase in reaction time is not present however, if the stimulus is strong. Other results from this research include the reaction time as function of stimuli/reaction location [Hasbroucq *et al.* 1986] and as a function of stimuli intensity [Hasbroucq *et al.* 1989]. The results are that the reaction is faster when the stimulus is strong, and when the reaction comes from the same location as the stimuli. The reaction times are generally between 200 and 500 mS. The potentially difficult choice of haptic feedback to the performer can be simplified by studying the physical reaction literature.

The reaction time literature can also be of use when designing the real-time interface between the performer and the synthetic musical instrument. More research is needed, however, before enough conclusions can be made. This issue is not further pursued, since the real-time issue is not investigated in this work.

[Friberg 1991] and [Friberg *et al.* 1991] introduced rules for the improvement of computer performance, which can give information on the most important expression parameters.

The control of a musical instrument is intimately related to the structure of the instrument and the production of the sound. Some good textbooks on the acoustics of musical instruments are [Backus 1970] and [Benade 1990] and [Fletcher *et al.* 1993].

2.3. Timbre Dimensions

Timbre is defined in [ASA 1960] as that which distinguishes two sounds with the same pitch, loudness and duration. This definition defines what timbre is not, not what it is.

Timbre is generally assumed to be multidimensional. For the sake of simplicity, it is assumed in this work that timbre is the perceived quality of a sound, where some of the dimensions of the timbre, such as pitch, loudness and duration, are well understood, and others, including the spectral envelope, time envelope, etc., are still under debate. In most research, however, the pitch, loudness and duration are dissociated from the timbre.

In general, it is accepted that the frequency/perceived pitch scale, or amplitude/ perceived loudness scale, is not linear [Handel 1989]. It is interesting to model the perceptive scale, since the values of the model would have a more intuitive scale, and the errors in the modeling would be perceptually minimized. For some parameters, such as the pitch, this effect is not modeled here, since there already exists an accepted musical scale, the 12 tones per octave scale.

Future work which models non-harmonic, non-acoustic instruments could potentially have much use of the frequency/perceived pitch and the amplitude/perceived loudness scales.

In this work, it is assumed that timbre models two different aspect of the sounds: The identity of the sound and the expression of the sound.

The identity of a sound is the ability to recognize a sound as the sound of, for instance, a piano, and the expression of a sound is the ability to recognize the sound as a high-pitched piano, or a soft piano, for instance.

Here, a survey of literature on timbre is presented. The conclusion of this survey will help in designing the models of the timbre.

2.3.1. Identity

The identity of a sound is defined in this work as the timbre cues that make possible the identification of the instrument that produces the sound. Other identities could define the

player of the instrument that produced the sound, the location of the instrument or the media that distributed the sound.

The difficulty of timbre identity research is often increased by the fact that many timbre parameters are more similar for different instrument sounds with the same pitch, than for sounds from the same instrument with different pitch. For instance, many timbre parameters of a high pitched piano sound are closer to the parameters of a high-pitched flute sound than to a low-pitched piano sound. Nevertheless, human perception always identifies the instrument correctly.

2.3.2. Pitch, Loudness and Duration

Pitch, loudness and duration are the most common expression parameters used for isolated sounds in music. Pitch defines the perceived note of the sound, loudness the perceived intensity of the sound and duration the length of the sound [Lindsay | 1977].

Pitch is in its simplest form seen as the fundamental frequency; this is the model adopted here. When the fundamental frequency is missing, it can be recreated from the difference of higher harmonic overtones.

Intensity is most often expressed in dB, sometimes in perceived dB, which is called phon, where the intensity at a given frequency is the same as the intensity at 1kHz. The sound also has an auditory threshold, under which it can no longer be perceived, and a pain threshold. Additionally, the dB scale can be converted to the loudness scale in sones. This scale indicates that the same change in dB doesn’t give the same perceived change in sones in low intensities as in high intensities. See [Handel 1989] for more details. The intensity is measured in linear scale throughout this document.

Duration is here expressed in milliseconds (mS); it is the length of the sound. No attempt has been made to find the perceived duration although it is believed that this work finds attack onsets close to the perceived onset. See [Gordon 1987] for a study of the perceptual attack time.

Research which aim is to understand the basic mechanism in hearing has been pursued for many years [Møller 1973]. This has given rise to more elaborate models, which take into account the functioning of the auditory system [Meddis *et al.* 1991a].

2.3.3. Dissimilarity Tests

The dissimilarity test is a common method of finding similarity in the timbre of different musical instruments. The dissimilarity tests are performed by asking subjects to judge the dissimilarity of a number of sounds. A multidimensional scaling is then used on the scores, and the resulting dimensions are analyzed to find the relevant timbre quality. [Grey 1977] found the most important timbre dimension to be the spectral envelope. Furthermore, the attack-decay behavior and synchronicity were found important, as were the spectral fluctuation in time and the presence or not of high frequency energy preceding the attack.

[Iverson *et al.* 1993] tried to isolate the effect of the attack from the steady state effect. The surprising conclusion was that the attack contained all the important features, such as the spectral envelope, but also that the attack characteristics were present in the steady state. Later studies [Krimphoff *et al.* 1994], refined the analysis, and found the most important timbre dimensions to be brightness, attack time, and the spectral fine structure.

[Grey *et al.* 1978], [Iverson *et al.* 1993] and [Krimphoff *et al.* 1994] compared the subject ratings with calculated attributes from the spectrum. [Grey *et al.* 1978] found that the centroid of the bark [Sekey *et al.* 1984] domain spectral envelope correlated with the first axis of the analysis. [Iverson *et al.* 1993] also found that the centroid of the spectral envelope, here calculated in the linear frequency domain, correlated with the first dimension. [Krimphoff *et al.* 1994] also found the brightness to correlate well with the most important dimension of the timbre. In addition, they found the log of the rise time (attack time) to correlate with the second dimension of the timbre, and the irregularity of the spectral envelope to correlate with the third dimension of the timbre. [McAdams *et al.* 1995] further refined this hypothesis, substituting spectral irregularity with spectral flux.

The dissimilarity tests performed so far do not indicate any noise perception. [Grey 1977] introduced the high frequency noise preceding the attack as an important attribute, but it was later discarded in [Iverson *et al.* 1993]. This might be explained by the fact that no noisy sounds were included in the test sounds. [McAdams *et al.* 1995] promises a study with a larger set of test sounds. It might also be explained by the fact that the most common analysis methods doesn’t permit the analysis of noise, which then cannot be correlated with the ratings.

2.3.4. Verbal Attributes

Timbre is best defined in the human community outside the scientific sphere by its verbal attributes. [von Bismarck 1974a] had subjects rate speech, musical sounds and artificial sounds on 30 verbal attributes. He then did a multidimensional scaling on the result, and found 4 axes, the first associated with the verbal attribute pair dull-sharp, the second compact-scattered, the third full-empty and the fourth colorful-colorless. The dull-sharp axis was further found to be determined by the frequency position of the overall energy concentration of the spectrum. The compact-scattered axis was determined by the tone/noise character of the sound. The other two axes were not attributed to any specific quality.

2.3.5. Noise

The noise of a musical instrument, or of any sound, is in itself a multidimensional attribute. Much work on the noise of the human voice has been done. [Richard 1994] offers a survey of speech noises. [Klingholz 1987] divides the aperiodic component into 2 types. The first type consists of the additive noises, which are colored or white noise, and not correlated with the pitched sound. Additive noises are either transients, or quasi-stationary. The other noise component is the random fluctuation of the fundamental frequency, jitter, and the random fluctuation of the amplitude, shimmer. Still another noise type is the change of waveform, which [Klingholz 1987] calls structural noise, but which is generally called aperiodicity.

For musical instruments, noise can be divided into additive noises, jitter, shimmer, and aperiodicity [McIntyre *et al.* 1981].

2.3.6. Roughness

Another important timbre attribute is roughness [Terhardt 1974]. Roughness is a measure of fast beats between two partials of the sound, which have the perceptual quality roughness. It is closely related to dissonance-consonance [Plomb *et al.* 1965]. The roughness, or dissonance, is most often used in the analysis of the consonance of two or more sounds, but it is equally applicable in the analysis of the roughness of one sound.

Roughness is related to the theory of critical bands [Zwicker *et al.* 1957], in that the partials that create the beat must be in the same critical band. Therefore, roughness is assumed to be zero in a harmonic sound with a fundamental frequency above 262 Hz [Terhardt 1974]. Roughness is not used in this work, although it seems promising in the

modeling of the transient of for instance the clarinet, where spurious frequencies sometimes increase the perceived roughness in the attack.

2.4. Additive Model

The additive model has been chosen in this work for the known analysis/synthesis qualities of this model. Many analysis/synthesis systems using the additive model exists today, including SMS [Serra *et al.* 1990], the lemur program [Fitz *et al.* 1996] and the diphone program [Rodet *et al.* 1997]. Other methods investigated include the physical models [Jaffe et al. 1983], the granular synthesis [Truax 1994], and the wavelet analysis/ synthesis [Kronland-Martinet 1988].

The additive model is well suited for the analysis of pitched sounds. In this model, the sound is supposed to be the sum of a number of sinusoidals with time-varying amplitude and frequency,

$$sound(t) = \sum_{k=1}^{N} a_k(t) * \sin\int_{\tau=0}^{t}(\omega_k(\tau)\partial\tau + \phi_{0,k}) \qquad \textbf{(2.1)}$$

The sinusoidals are denoted partials which corresponds to harmonic overtones when the sound is harmonic. Then the frequencies of the partials are multiples of the fundamental frequency.

The frequency of the harmonic partials is equidistant in the frequency domain. The first many harmonic overtone frequencies fall close to the notes in the 12-tone/octave scale. The relation between the strong overtones of compound musical sounds is what defines the consonance of the interval [Kameoka *et al.* 1969].

The additive parameters are best viewed in a three-dimensional plot, as shown in figure 2.1, where the axes are time, frequency, and amplitude.

The lines in the plot indicate the evolution of the amplitude and frequency of each partial. This plot shows a test signal which is harmonic with a fundamental frequency of 100 Hz.

All frequencies are static and the partial frequencies are 100, 200, 300, 400, 500, 600, 700 and 800 Hz.

The closest line (to the left) is the fundamental. The amplitude of the fundamental is first zero for 100 mS, then it follows a linear slope from 1500 to 500 for 800 mS and then it is zero for another 100 mS. The amplitude of the seven upper partials is half of the amplitude of the preceding partial. The total duration of the sound is 1 second.

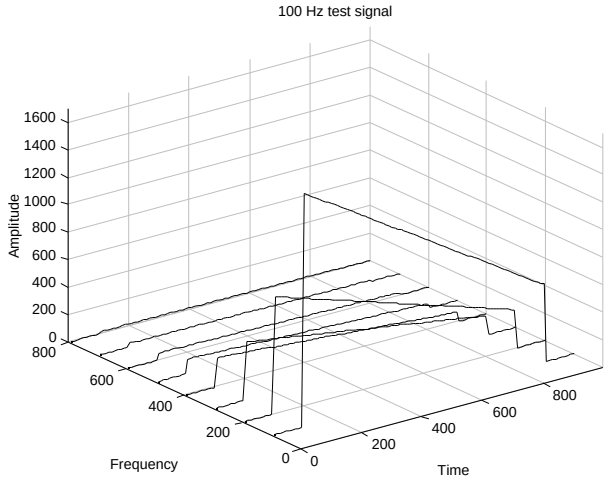


Figure 2.1. Additive parameters plot. The x axis is time in mS, the y axis is frequency in Hz and the z axis is amplitude.

2.4.1. Time-Frequency Analysis

The additive parameters are found by a time/frequency analysis. In the time/frequency analysis, the amplitudes and frequencies are estimated at each time step. A time resolution and a frequency resolution are involved in the time/frequency analysis. Rather than talk about frequency resolution, frequency discrimination is often a more valid criterion. Unfortunately, time resolution and frequency discrimination are mutually incompatible, which means that if a better time resolution is sought, then a worse frequency discrimination is obtained. In general terms, a better time resolution is obtained for higher fundamental frequencies of harmonic sound, which is in accordance both with the fact that the higher frequencies generally have faster attack times (see the analysis of the IDA model parameters in Chapter 9), and that frequency spacing is larger for these sounds. The time resolution should be at least as good as the fastest transient time under analysis, in the order of a few mS.

2.4.2. Phase

There have been many debates on the importance of the relative phase of the sinusoidals. The survey of the literature is not facilitated by the confusion of initial phase and running phase (beats). Only the initial phase are studied here. This corresponds to $\phi_{0,k}$ in equation (2.1). Early research on the functioning of the ear had two opposing views, the frequency domain model, which states that phase differences cannot be heard, and the temporal model, which states that phase is important.

Perceptive experiments, cited below, involving two, three or more sinusoidals are formal. The phase is important. [Plomb *et al.* 1969] resumes the previous research, and performs additional experiments. His conclusion is that phase difference can be heard, and he further compares the maximum effect of phase change to the perceptual difference of three close vowels. He also concludes that the phase effect is greater for low frequencies. [Buunen 1976] uses the phase to compare envelope detection and finds that envelope detection in the human can be described as a low-pass filter with a cut-off frequency of between 30 and 100 Hz. This translates into a better envelope detection if the envelope is slow, or if the envelope change is large.

[Paterson 1987] makes additional experiments and further models phase sensitivity and [Meddis *et al.* 1991b] offer a refined model of the auditory system, which explains at least some of the phase effects. This model replaces the early temporal peak-picking methods for fundamental frequency estimations with a series of autocorrelations of band-pass filtered signals. The argument is that the ear is mostly phase sensitive only within frequency channels. Paterson experiments involve phase sensitivity as a function of fundamental frequency, harmonic number, level, and duration. His conclusions are “a) the timbre of musical notes below middle C on the keyboard depends on component phase relations, and b) the quality of most mens’ voices and many womens’ voices depends on component phase relations.” [McAuley *et al.* 1986] seems to reach the same conclusions in their work on analysis/synthesis using additive parameters.

Although the initial phase is important to the perception of a sound, “this effect is quite weak, and it is generally inaudible in a normally reverberant room where phase relations are smeared” [Risset *et al.* 1982].

In conclusion, the initial phase seems important for timbre perception in low frequencies (below middle C, 262 Hz), at least in a non-reverberant listening situation. Unfortunately, neither the initial phase, nor phase coupling, has been modeled in this thesis. It is therefore labeled future work.

2.5. Database

To have some material to analyze, it is necessary to have a database of sounds. Several such databases are available on the commercial marked; the most widespread is probably the McGill University Master Samples (MUMS) [Opolko *et al.* 1988].

Commercial musical instrument databases do not generally have different tempi, intensity or style for the full playing range of a musical instrument. New recordings were therefore judged necessary.

Based on the preliminary research in timbre and control, a selection of different musical instruments from different families has been recorded. The facilities and material can be called semi-professional, all recordings being done on DAT and transferred digitally to the computer network. Some of the performing musicians were professional and some were amateurs. This doesn't seem to influence the quality of the recordings much, since the material is essentially non-musical.

The instruments in the database are the violin, the viola, the cello, the saxophone, the clarinet, the flute, the soprano voice and the piano. Some of the instruments, such as the violin, have many degrees of physical freedom; the speed, force angle and direction of the bow is only a small subset. Others instruments only have a few degrees of physical freedom; the piano player, for instance, can influence only the position, or the speed, of the key(s), and the pedals.

The recording details can be found in appendix A.

2.6. Conclusions

The sound of the musical instrument can be qualified by the timbre or the identity and the gestures. Gestures associated with musical instruments are well defined by common musical terms, such as note, loudness, tempo or style. Timbre defines the identity and the expression of a musical sound. It seems to be a multi-dimensional quality. Generally, timbre is separated from the expression attributes pitch, loudness, and length of a sound. Furthermore, research has shown that timbre consists of the spectral envelope, an amplitude envelope function, which can be attack, decay, or more generally, the irregularity of the amplitude of the partials, and noise. Other perceptive attributes, such as brightness and roughness, can also be helpful in understanding the dimensions of timbre.

The quasi-harmonic musical instrument sounds are generally well defined by their additive coefficients, which, in a listening situation without reverberation, should retain the phase relations if the fundamental frequency is below middle C (262 Hz).

Chapter Three

3. Fundamental Frequency Estimation

In this chapter the estimation of the fundamental frequency of a musical sound is presented. The fundamental frequency is generally seen as the frequency of the first strong partial (the fundamental), or as the frequency difference between two adjoining harmonic overtones. The frequency differences are used to find the fundamental frequency here and the estimation of the fundamental frequency of quasi-harmonic sounds is improved in this work by fitting the estimated frequencies to the ideal quasi-harmonic frequencies. A fundamental frequency tracker is also introduced. Furthermore, an estimation of strong frequencies present in a musical sound is presented. The strong frequency estimations found in this chapter are used in the time/frequency analysis in the next chapter.

3.1. Introduction

The fundamental frequency of a musical sound is an important timbre attribute. The fundamental frequency is here found by matching a stretched harmonic curve to the frequencies of the partials found by the Fast Fourier Transform (FFT) analysis. Not all stretched harmonic components are found by the initial FFT analysis. Those not found are

reinserted, and the non-harmonic partials are removed before the curve fitting. The frequencies extracted from the stretched curve along with the strong non-harmonic components are used as the basis for the estimation of the time-varying frequency and amplitude of the partials.

Several algorithms for the estimation of fundamental frequency have been presented in the last few decades. The fundamental frequency estimation can be done in the time domain [Rabiner *et al.* 1976], [Rabiner 1977], [Kroon *et al.* 1990], the cepstrum domain [Noll 1967], or the frequency domain [Doval *et al.* 1991]. [Freed *et al.* 1997] proposes a database of a wide range of sounds for the objective comparison of pitch estimation techniques.

The frequency domain estimation of the fundamental frequency seems to be predominant today, and an implementation of a frequency domain fundamental frequency estimator is presented here. The general idea is to estimate the fundamental by the difference in frequency of the neighboring harmonic components. This standard method for the estimation of fundamental frequency is improved in this work by matching a perfect stretched harmonic curve to the estimated quasi-harmonic partial frequencies.

This chapter starts with the estimation of the FFT candidates in section 3.2, the fundamental frequency estimation is presented in section 3.3, and the quasi-harmonic frequencies are estimated in 3.4, along with non-harmonic components, which are here called the spurious frequencies. The pitch tracker is presented in section 3.5, and the chapter ends with a conclusion.

3.2. FFT candidates

The FFT candidates are found by performing an FFT on a strong segment of the sound, and estimating the frequencies and amplitudes of the peaks of the absolute of the FFT. Weak peaks close to stronger peaks are removed by a line that imitates the masking of the auditory system. Although the sounds are supposed to be pseudo-harmonic, no such hypothesis is used in the FFT analysis. All candidates that are strong enough are saved. The frequency and amplitude estimation is improved by interpolating between frequency bins. More details on the FFT can be found in, for instance, [Steiglitz 1996] and [Press *et al.* 1997].

3.2.1. Frequency and Amplitude Estimation

The estimation of strong partials is done through the Fast Fourier Transform (FFT) on a strong segment of the sound. The strong segment is defined as being the segment after the strongest segment in the sound. This is usually the segment after the attack segment. This segment is used, since there is often too much transient behavior in the attack segment.

The FFT is a fast implementation of the discrete Fourier transform,

$$y_n = \sum_{k=0}^{N-1} s_k e^{i 2\pi n k / N} \quad \textbf{(3.1)}$$

where s_k is the discrete time signal and *n* is the frequency bin index, from which the frequency can be calculated,

$$f_k = s_r n / N \quad \textbf{(3.2)}$$

s_r is the sample rate. The inverse discrete Fourier Transform is defined as,

$$s_n = \frac{1}{N} \sum_{k=0}^{N-1} y_k e^{-i 2\pi n k / N} \quad \textbf{(3.3)}$$

In general, the time signal is multiplied by a window to avoid discontinuity effects,

$$y_k = FFT(s_k \; h_w) \quad \textbf{(3.4)}$$

In this work the window used is a hamming window [Harris 1978],

$$h_w = 0.54 - 0.46\cos(2\pi k / (N-1)) \quad \textbf{(3.5)}$$

When the frequency domain signal y_k is available, the frequencies and amplitudes can be found simply by looking for maximums of the absolute value of y_k. When a maximum is found in i_y then,

$$f_k = s_r i_y / N \quad \textbf{(3.6)}$$

and,

$$a_k = \left| y_k(i_y) \right| \quad \textbf{(3.7)}$$

As can be seen, the frequency resolution is dependent on the blocksize N. A better frequency resolution can be obtained by interpolation if a gauss window is used,

$$h_w = e^{-\frac{(k-\frac{N}{2})^2}{2\sigma^2}} \quad \textbf{(3.8)}$$

Then it can be shown that, if we know the maximum in the FFT domain, i_y, the maximum is displaced by,

$$cor = \frac{0.5*(\log(|y_k(i_y-1)|)-\log(|y_k(i_y+1)|))}{(\log(|y_k(i_y-1)|)-2.0*\log(|y_k(i_y)|)+\log(|y_k(i_y+1)|))} \quad \textbf{(3.9)}$$

and the new frequencies and amplitudes are,

$$f_k = \frac{s_r(i_y+cor)}{N} \quad \textbf{(3.10)}$$

$$a_k = \exp(\log(|y_k(i_y)|)-0.25*cor*(\log(|y_k(i_y-1)|)-\log(|y_k(i_y+1)|)) \quad \textbf{(3.11)}$$

This interpolation is helpful, even if a gauss window is not used. Initial comparisons indicate that the frequency estimation is improved by the same order of magnitude as using two FFTs one sample apart and calculating the frequency from the phase differences. Other methods of decreasing the errors of the frequency estimation can be found in [Quinn 1994].

When a maximum is found, the frequency domain vector y_k is set to zero below i_y while the derivative is positive, and above i_y when the derivative is negative. The search for maximums continues until more than *M* partials have been found, or until the partial is weaker than a ratio times the strongest partial.

3.2.2. Masking

In order not to get too many unusable partials, here called spurious partials, which are usually found close to the quasi-harmonic partials, y_k is superposed by a window w_y, which is 0.9 multiplied with the maximum of y_k over $2*w_{sz}$ samples. This puts a line slightly below the maximum of the partials, but above the noise and most of the spurious partials. The FFT-based peak search is illustrated in figure 3.1 for a piano sound. The x-axis is the frequency and the y-axis is the log of the amplitude.

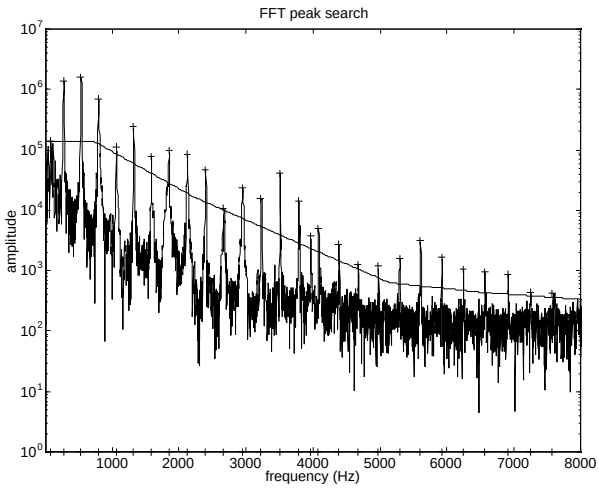


Figure 3.1. FFT-based peak search for a piano sound. Found peaks are marked with a '+'. The solid line below the peaks is the masking line.

The plus signs denote the amplitudes and frequencies of the partials found. The spurious frequencies and noise are generally placed below the masking line. The line imitates the auditory masking of weak partials [Small 1959], [Schroeder *et al.* 1979]; however, the goal is not to estimate only perceived partials, but to eliminate noise, since weak partials can become perceptible by some subsequent processing of the data.

Unfortunately, the masking sometimes leaves some undesired spurious partials in the analysis.

The FFT candidates for a piano sound can be seen in figure 3.2. The x-axis is the frequency, and the y-axis is the amplitude. The strong, harmonic partials of the sound are easy to see above the noise floor. The weak partials below strong partials are generally spurious partials, or sometimes very weak harmonic partials in between stronger partials.

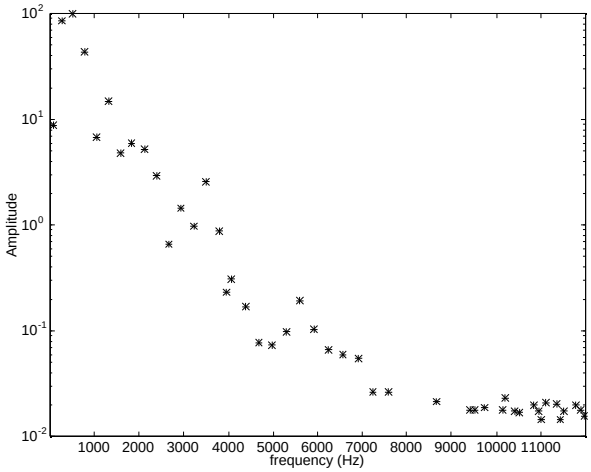


Figure 3.2. FFT candidates for the piano sound.

The high frequency partials seem to be close to the noise floor, although many of them have the correct frequency.

The next sections indicate the method developed in this work to find the fundamental frequency, the harmonic components, and other non-harmonic partials that are strong enough to be perceived (spurious partials).

3.3. Fundamental frequency estimation

The initial frequency candidates are here used to estimate the fundamental frequency. The process is as follows. First, only the frequencies whose amplitude is above a certain threshold are used. Next, the frequency differences are calculated, using the first frequency as the first difference. Then, all frequency differences that lie outside a percentage of the mean frequency are removed. The percentage is lowered and the process is repeated until the percentage is low enough. The mean of the filtered frequencies is the first estimation of the fundamental frequency. This estimation is used to add missing harmonic frequencies and remove non-harmonic frequencies from the FFT frequency candidates. The resulting frequencies are now the overtones of a quasi-harmonic sound. By quasi-harmonic is meant

that the frequencies can be either stretched, or compressed, so that the frequency of the harmonic partial *k* is a little higher or lower than *k* times the fundamental.

3.3.1. Frequency Difference Fundamental Estimation

The first step of the process is to calculate the frequency differences of the FFT frequency candidates,

$$fd = f_1 - 0, f_2 - f_1, f_3 - f_2, \ldots \quad \textbf{(3.12)}$$

Then, the mean of the frequency difference is calculated, which is the first fundamental frequency estimation,

$$fund = mean(fd) \quad \textbf{(3.13)}$$

Now frequency differences whose values differ by a threshold from the mean of the frequency differences are eliminated. This process is repeated with smaller and smaller threshold, until no more frequencies are eliminated.

It is necessary to take into account the inharmonicity of the sound, since the frequency difference of higher partials of, for instance, the piano can be very different from the fundamental.

This is done using the difference of the frequency difference, which is calculated for adjoining harmonic partials as,

$$fdd_k = fd_k - fd_{k-1} \quad \textbf{(3.14)}$$

and removing the local average of the difference of the frequency difference from the frequency differences,

$$fd'_k = fd_k - \frac{1}{L}\sum_{l=1}^{L} fdd_{k-l} \quad \textbf{(3.15)}$$

L is in the order of a few overtones.

The frequency differences for the frequencies whose amplitude is above a certain threshold are shown in figure 3.3 (top). The frequencies which lies within a threshold of the mean of the difference of the frequencies are shown in figure 3.3 (bottom). Remember that they are corrected for inharmonicity.

The mean of the filtered frequencies in figure 3.3 (bottom) is the first estimation of the fundamental frequency. The estimation is 265.5 Hz.

The frequencies that are removed are often at double the fundamental frequency if a harmonic partial is missing. These can be divided by two to obtain a better fundamental frequency estimation.

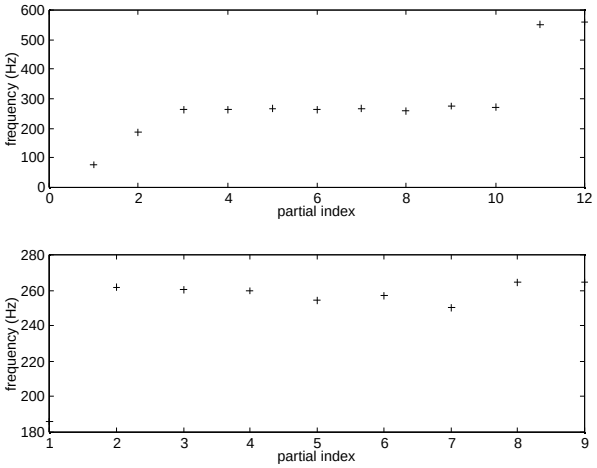


Figure 3.3. Frequency differences for the piano before (top) and after (bottom) filtering.

The fundamental frequency estimation at this point is,

$$fund = mean(fd') \qquad \textbf{(3.16)}$$

where *fd'* is the frequency differences vector of length *N* after elimination of linear inharmonicity deviations.

3.3.2. Missing Frequencies

With the fundamental frequency estimation, *fund*, it is possible to recognize the frequency candidates that are indeed harmonic components, and eliminate the non-harmonic components.

It is also necessary to add missing harmonic components in order to perform the curve fitting on the stretched harmonic curve. This is done by estimating a local fundamental frequency for each overtone as shown in equation (3.12), and adding fd_k to the preceding harmonic partial if no harmonic component is found. To reduce the error, the local fundamental is averaged over several partials.

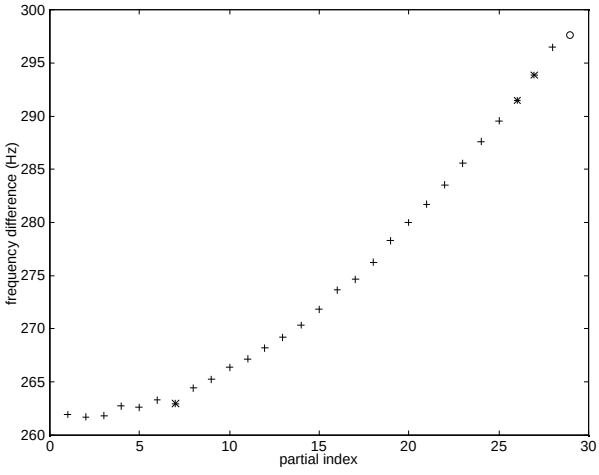


Figure 3.4. Frequency differences after cleaning frequencies. '+' indicates original frequencies, '*' are chosen from several candidates, and 'o' are new inserted frequencies.

In figure 3.4 the result of the cleaning of the frequencies of a piano sound is shown. The frequency differences, i.e. the frequencies of the partials minus the frequencies of the preceding partials, are plotted divided by the partial index.

The inharmonicity, that is, the stretched frequency of the upper partials, see section 3.3.3, is clear to see. Only few frequencies are inserted, these are denoted by a 'o'. The other partial frequencies have been copied from the analyzed frequencies, denoted with '+' signs, or chosen from several candidates, denoted with '*'.

The fundamental frequency can here be estimated by eye to about 262 Hz.

3.3.3. Fit Stretched Harmonic Curve.

When the fundamental frequency is found, the harmonic overtones of many musical instruments can be analyzed, simply by looking at multiples of the fundamental frequency. Unfortunately, not all musical instruments have pure harmonic partial frequencies. For instance, the piano, due to the stiffness of the strings, has sharp upper partial frequencies; i.e. the frequencies are higher than the fundamental multiplied by the harmonic partial index. According to [Fletcher 1964], "the 40th partial can be two full notes sharp". The frequencies that are not exactly harmonic are said to be quasi-harmonic. The formula for the quasi-harmonic frequencies of a stiff piano string is,

$$f_k = k f_0 \sqrt{1+\beta k^2} \qquad \textbf{(3.17)}$$

where f_0 is the fundamental frequency and β is the inharmonicity.

The values of f_0 and β are found using a nonlinear least-squares curve fit [Moré 1977]. To minimize the error in the important low partials, the curve fit is done on the frequencies divided by the partial index. The frequencies divided by the partial index for the piano sound are plotted in figure 3.5 with the estimated frequencies given by the formula (3.17).

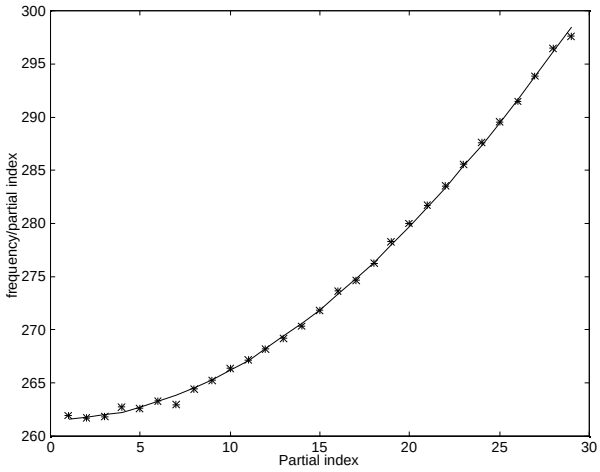


Figure 3.5. Frequency divided by the partial index with estimated stretched curve for the piano sound.

The frequencies in figure 3.5 are not the same as the frequencies in figure 3.4, which are the difference between adjoining partials.

The inharmonicity of the piano is clearly seen in figure 3.5. The fundamental frequency, which is the frequency of the stretched harmonic curve at index 1, is calculated to be 261.5 Hz. The inharmonicity index β is $3.6*10^{-4}$.

With the stretched harmonic curve fit, the estimation of the fundamental frequency is terminated.

3.4. Initial Frequencies

In this paragraph, the strong partials of a sound are found. These are a good initial estimation of the frequency content of a sound, and they can be used to further analyze the sound. One obvious way of finding these frequencies is by using the FFT candidates as described in section 3.2. These are often missing some harmonic partials, and introduce too many spurious frequencies or false partials. For these reasons, the frequencies of the stretched curve found in paragraph 3.3.3 are used instead. The sounds are supposed to be quasi-harmonics, but there can also be strong non-harmonic partials, which are here called spurious frequencies. These are therefore also added to the initial frequencies.

3.4.1. Harmonic Frequencies

The harmonic frequencies are found from the FFT candidates as explained in section 3.3 by first estimating the fundamental frequency, and then removing spurious frequencies, adding missing harmonic components, and finally fitting a stretched harmonic curve to these frequencies, which ensures that large frequency estimation errors are eliminated.

3.4.2. Spurious Partials

The stretched harmonic partials found in the preceding paragraph are often enough to define a sound. Sometimes, however, different behavior introduces non-harmonic partials [Conklin 1997]. These are here called spurious partials, and they can sometimes be stronger than the neighboring harmonic frequencies. It is therefore necessary to introduce them in the initial frequencies. The spurious frequencies can also participate in the identification of the instrument, although they are rarely desired in a musical situation.

The spurious frequencies are found by comparing the original FFT candidates from section 3.2 with the stretched inharmonic frequencies found in paragraph 3.3.3. A spurious

frequency is introduced, if it is sufficiently far away from the neighboring frequencies and if it is relatively strong compared with the neighboring frequencies and compared to the strongest partial.

The combination of the quasi-harmonic frequencies and the spurious frequencies are used as initial frequencies in the linear time/ frequency analysis in the next chapter.

3.4.3. Spectrogram Analysis

The spectrogram is a good starting point for the estimation of the frequency content of a sound. Unfortunately, the spectrogram is as noisy as the FFT used in the creation of the spectrogram. However, image-processing techniques can be useful in the analysis of spectrograms. The estimation of the initial moving frequency could be improved by the image analysis methods used in the scale-space research [Lindeberg 1996]. Notably, Joachim Weichert has introduced an anisotropic diffusion filtering in [Weickert 1999] and also performed some initial experiments on the spectrogram of musical sounds in [Weickert 1998]. This topic has not been further pursued here, but it seems promising, if initial frequencies are necessary, as is the case for the LTF additive analysis in the next chapter.

3.5. Pitch Tracker

The pitch is one of the most important timbre attributes, and it is also the most common expression control in musical instruments. Therefore, a pitch tracker is necessary if the evolution of all the timbre attributes is to be followed. The pitch track is done in three steps, first the fundamental frequency of each short segment is found using the methods presented in section 3.3. Then the instantaneous frequency [Boashash 1992] is found by removing all but the fundamental frequency from the FFT of the segment, and doing an inverse FFT on the result. Finally, the frequency evolution is segmented. This work is only a feasibility analysis, and it is not used in the rest of this thesis. More proved methods for pitch tracking can be found in, for instance, [Medan *et al.* 1991] for the speech signal, and in [Quirós *et al.* 1994] and [Dorkan *et al.* 1994] for musical signals.

This feasibility study is presented using two short melodies, a flute melody, and a viola melody, the spectrogram of which are shown in figure 3.6 and figure 3.7. Notice the absence of the fundamental in the two first notes of the viola melody in the spectrogram plot in figure 3.7.

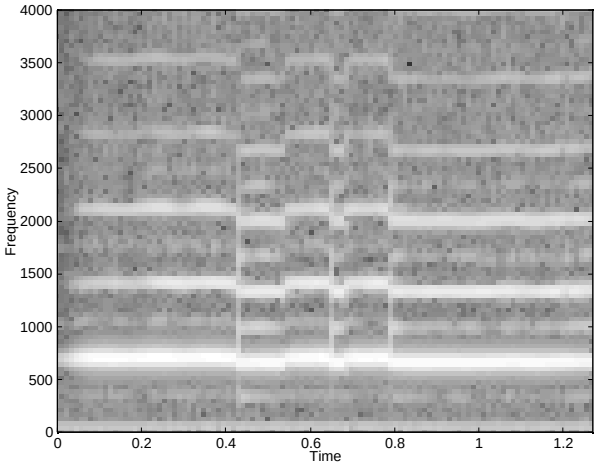


Figure 3.6. Spectrogram of the flute melody.

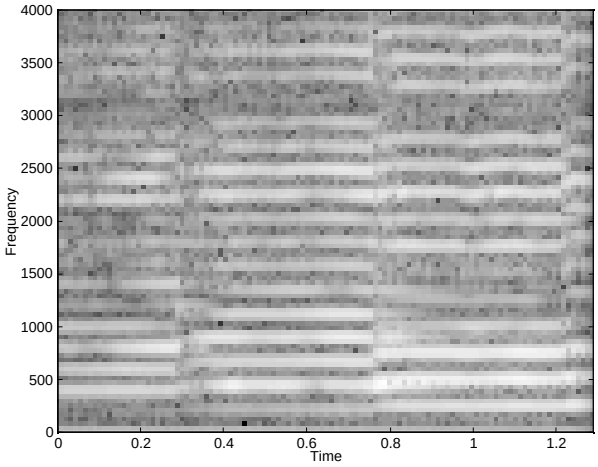


Figure 3.7. Spectrogram of viola melody.

3.5.1. Moving Fundamental Frequency

The moving fundamental frequency is found by the method used in section 3.3, using a fairly short blocksize. This fundamental frequency estimation, while relatively exact, is unfortunately rather slow. The fundamental frequency estimation of a short flute extract is shown in figure 3.8 and for a short sound of four notes of a viola is shown in figure 3.9.

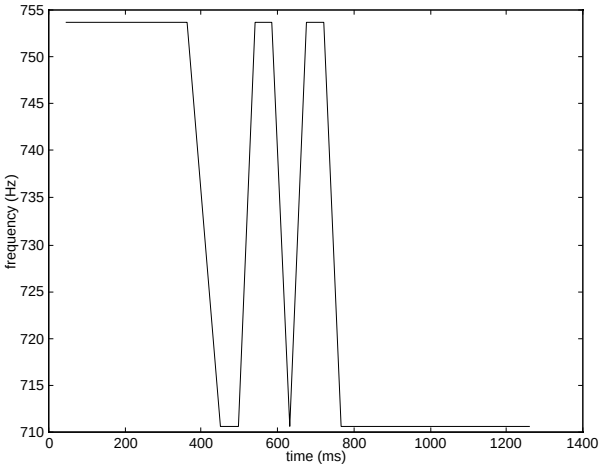


Figure 3.8. Moving fundamental frequency for the flute melody.

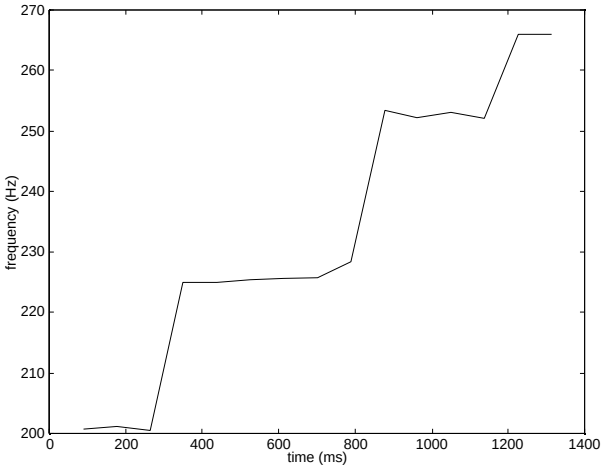


Figure 3.9. Moving fundamental frequency for the viola melody.

The estimation could be significantly improved, if the blocks were to be aligned at the pitch change times, or, if an overlap analysis were performed.

3.5.2. Instantaneous Frequency

The moving fundamental frequency estimation found in paragraph 3.5.1 sometimes doesn't have a good enough time resolution. Therefore, a better estimation is found by removing all but the fundamental of the FFT, and doing an inverse FFT on the result,

$$\hat{y} = FFT^{-1}(FFT(snd)\ \ h_{fund}) \qquad \textbf{(3.18)}$$

where h_{fund} is a window that covers the fundamental only in the frequency domain.

The moving frequency is then the derivative of the arc tangent of the result of the inverse FFT,

$$fr = \frac{s_r}{2\pi}\frac{\partial}{\partial t}\left(\arctan\left(\frac{\Im(\hat{y})}{\Re(\hat{y})}\right)\right) \qquad \textbf{(3.19)}$$

To reduce the amount of data, the mean of each period of *fr* is taken. The resulting frequency is shown in figure 3.10 and figure 3.11. Notice how badly this method works for the viola sound. This is explained by the fact that the fundamental frequency is almost non-existent for the first two notes.

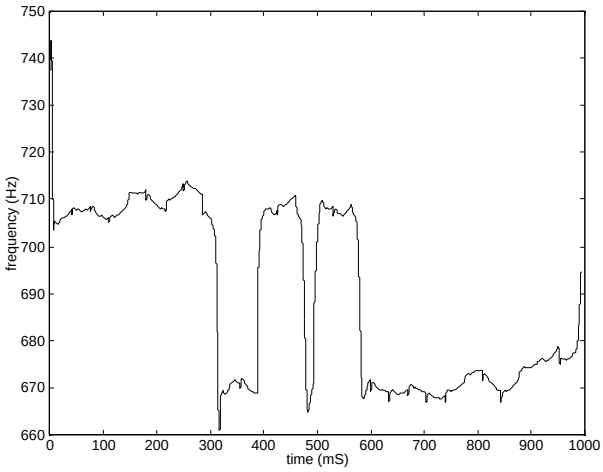


Figure 3.10. Instantaneous frequency of the flute melody.

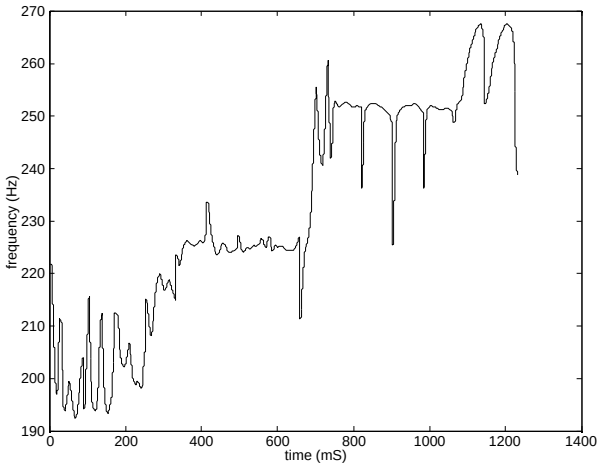


Figure 3.11. Instantaneous frequency of the viola melody.

3.5.3. Curve Segmentation

The instantaneous frequencies found in paragraph 3.5.2 are error-prone and not easily manipulated. The frequencies are therefore simplified into a curve, where short deviations are removed, and static parts are simplified into line segments.

The curve is found by first creating a coarser frequency curve by taking the mean over 256 samples. A new note is now found by moving the time ahead, and checking the frequency difference. When a big enough difference is found, and the time gap is big enough, a new note is found. The time is now reversed in the fine time resolution frequency curve until the old note is found. This is the start time for the new note.

The resulting frequency curves for the flute and viola melodies are shown in figure 3.12 and figure 3.13.

This method is very helpful in removing some of the noise on the instantaneous frequencies, without losing the good time resolution. No comparison has been made with other pitch tracking algorithms. Pitch tracking is a very difficult area, and it is believed that the method presented here, at this stage, is not stable enough. More work remains before an unsupervised use of the method can be undertaken.

The pitch track has been used on the analysis of vibrato sounds with some success. However, the pitch track is less stable than the fundamental frequency estimation in section 3.3, and in an automatic analysis situation, which is the case in this work where hundreds of sounds are analyzed, the pitch track is not yet stable enough. Since most sounds have static frequencies, it has not been a major concern and the pitch tracker has not been used in the rest of this work.

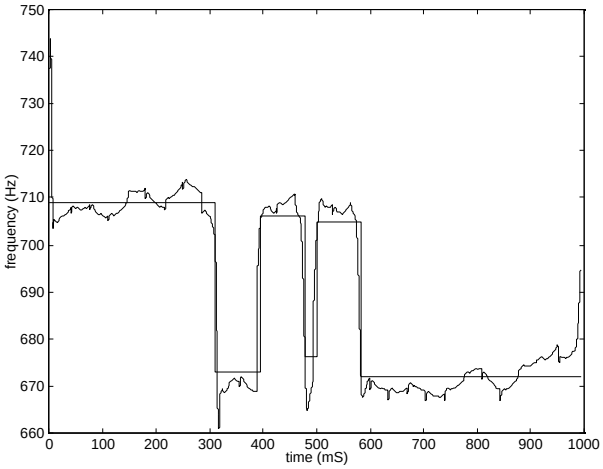


Figure 3.12. Instantaneous frequency and extracted curve for the flute melody.

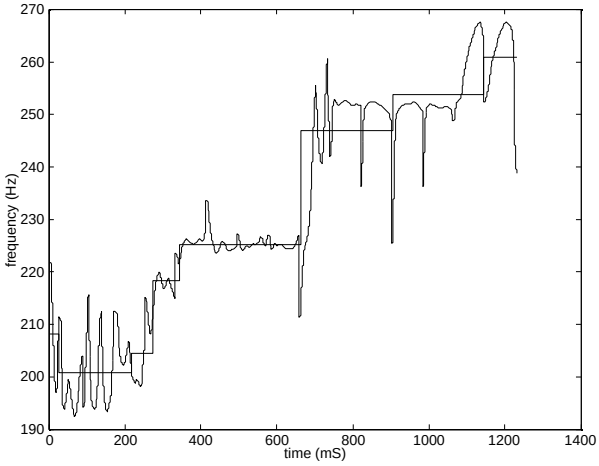


Figure 3.13. Instantaneous frequency and extracted curve for the viola melody.

3.6. Conclusions

The estimation of fundamental frequency, initial frequencies, and moving pitch was presented. This work has improved the classical fundamental frequency estimation by a curve fit with a stretched harmonic curve, which fits the frequency of the partials of stiff strings. This permits the estimation of the pitch of quasi-harmonic sounds, such as the piano tones.

The initial frequencies of a sound are here defined to be the harmonic overtones, with additional strong amplitude spurious partials.

The pitch track is performed using frequency domain filtering and inverse FFT, but the result is not yet satisfactory. Better results could potentially be obtained if image-processing techniques were used on spectrograms.

In conclusion, this chapter presents a good fundamental estimation, a fair initial frequency estimation, and a promising pitch tracking method. The initial frequencies are used in the linear time frequency analysis in the next chapter.

Chapter Four

4. Analysis/Synthesis

In this chapter two methods for analyzing musical sounds are compared. The additive model is used, where the sound is modeled as a sum of sinusoidals, also called partials, with time-varying amplitude and frequency. The sounds can be resynthesized with no loss of quality, if the analysis is good, by adding the sinusoidals together.

Two analysis methods are compared, the classical FFT-based method, and a new method, based on a linear time-frequency representation [Guillemain *et al.* 1996]. This method, which is here called the LTF analysis, is used in the following chapters due to the better time resolution of the analysis, which means that faster transients, such as the attack of the sound of the piano, can be analyzed more accurately. The LTF analysis is improved in this work by the estimation of initial frequencies in the preceding chapter, which permit a stable, unsupervised analysis of musical sounds.

4.1. Introduction

The analysis/synthesis of sounds of musical instruments is generally accomplished by using a model of a sum of sinusoidals. Here two techniques for the analysis of musical sounds are compared, the FFT-based analysis [McAuley *et al.* 1986], and the linear time/frequency analysis [Guillemain *et al.* 1996].

Already in the last century, musical instrument tones were divided into their Fourier series [Rayleigh 1896]. Early techniques for the time-varying analysis of the additive parameters are presented by [Matthews *et al.* 1961] and [Freedman 1967]. [Robinson 1982] gives a historical perspective of spectrum estimation methods. Other more recent techniques for the analysis of musical signals are the proven heterodyne filtering [Grey *et al.* 1977], the wavelet analysis [Kronland-Martinet 1988], the atomic decomposition [Chen *et al.* 1996], [Gribonval *et al.* 1996] and the modal distribution analysis [Pielemeier *et al.* 1996]. [Ding *et al.* 1997] has presented an interesting analysis by synthesis method.

Synthesis of the additive parameters has been done in real time for many years [Jensen 1989].

The analysis of musical signals is done in the time/frequency domain. There are two resolutions to the analysis, the time resolution, where a resolution of a few mS is necessary, and the frequency resolution, where an accuracy of a few cents is necessary [Pielemeier *et al.* 1996]. There are 100 cents between each semitone. Generally, not so much the frequency resolution is a problem, but instead the separation of partials in the frequency domain. The analysis is often a compromise between a good separation, and a good time-resolution. The FFT-based analysis generally optimizes the time-resolution by a two-pass analysis, one with a good time-resolution, and one with a good frequency-resolution. Nonetheless, it has a poor time-resolution, and several alternatives with better time resolution have been introduced to replace the FFT-based analysis.

This chapter starts with an implementation of the FFT-based analysis in section 4.2 and the linear time-frequency analysis is presented in its simplest form in section 4.3. A comparison is made between the two methods in section 4.4, the resynthesis is discussed in section 4.5, and finally a conclusion is proposed.

4.2. Fast Fourier Transform Based Additive Analysis

Several FFT-based [Allen 1977] sinusoidal analysis systems for sounds have been presented in the past [McAuley *et al.* 1986], sometimes with the addition of a stochastic component model of additive noise [Serra *et al.* 1990], [Møller 1996].

The FFT-based analysis is generally done on a sliding time-domain window. The FFT peaks are found by analyzing the FFT of a windowed time signal, as explained in the fundamental frequency estimation in Chapter 3. The peaks for a segment are then attached to the preceding segments' partial tracks.

4.2.1. Sliding Window Analysis

The model of the sound to analyze is a sum of sinusoidals with varying frequency and amplitude,

$$s(t) = \sum_{p=1}^{N} a_p(t)\sin(\int_{t=0}^{\tau} \omega_p(\tau)\partial\tau + \phi_{0,p}) \qquad \textbf{(4.1)}$$

The FFT is done on overlapping blocks of the signal *s(t)*, the blocksize is B and the block is *k*. The stepsize is B_s, so the peak searching for the block k is done on the samples of *s* from $k\,B_s$ to $k\,B_s+B$. The FFT is done using a hamming window. The output of the FFT is then,

$$y_k = FFT(s(kB_s ... kB_s + B - 1)\quad h_w) \qquad \textbf{(4.2)}$$

y_k is then used to look for peaks in the frequency domain as explained in Chapter 3. The window h_w is assumed to be a hamming window with normalized amplitude, so that the sum of all elements in h_w equals one [McAuley *et al.* 1986]. The output amplitudes and frequencies from block *k* are f^k and a^k. Each block can have a variable number of peaks.

4.2.2. Better Timing Resolution

The frequencies found above are used to perform a discrete fourier transform (DTF) on the exact frequency, using a window size *B* that is four times the period of the fundamental.

$$y_k(\omega) = \sum_{n=0}^{B-1} s(n + kB_s)e^{i\omega n}\quad h_w(n) \qquad \textbf{(4.3)}$$

This gives a slightly better amplitude value and the best timing-resolution with an FFT-based method obtained in this work. Shorter windows do not separate partials well enough.

[McAuley *et al.* 1986] used a hamming window with a size of 2.5 times the period, but it has not been possible to recreate their results here. [Serra *et al.* 1990] uses a Kaiser window of 4 period length and [Ando *et al.* 1993] uses a window size of four periods with a hanning window.

[Harris 1978] discussed the use of windows in harmonic analysis using the discrete Fourier transform.

4.2.3. Partial Track

In order to get a useful series of partials it is supposed that the frequencies and amplitudes can be connected in a series of connected lines, called tracks. The frequencies of these tracks can be harmonic, but they don't have to be, and there are often some shorter spurious partials in between the long strong (harmonic) tracks. Several methods for tracking partials have been developed, local optimized techniques, [McAuley *et al.* 1986], [Serra *et al.* 1990], or globally optimized techniques using hidden markov modeling [Depalle *et al.* 1993].

Here, a simple local optimized algorithm is used. When the frequencies and the amplitudes are slowly varying, and the sounds are harmonic, the task of connecting the points is fairly easy, but noise and natural variations often disturb the partials.

Supposing the partials up to time segment *k* have been connected. The *k* block has *N* partials and the *k*+1 block has *M* partials. Generally $M \neq N$.

The partials connect if the difference in frequency, and perhaps also the difference in amplitude, is small. All the close frequencies are analyzed and a matching value is calculated for each one of them,

$$match(n,m) = k_a\left|a_{k+1}^{n} - a_k^{m}\right| + k_f\left|f_{k+1}^{n} - f_k^{m}\right| \qquad \textbf{(4.4)}$$

where partial n from block *k*+1 is connected to the partial from block *k* with the best (lowest) match. The weights k_f and k_a are chosen experimentally. In this work, as is generally the case, k_f is set to one, and k_a is set to zero. A more stable tracking is obtained if the slopes of the frequency and amplitude are used. Notably, partial crossing is then possible [Depalle *et al.* 1993]. It is also worth noting that the tracking performs much better when the frequency and amplitude estimations are good. A notable improvement was observed when the spectral interpolation was used.

A track dies if no match is made for several blocks, and a track is born if there is no match possible for a partial.

4.2.4. FFT Conclusions.

The results of the FFT-based analysis can be seen in figure 4.1 to figure 4.4.

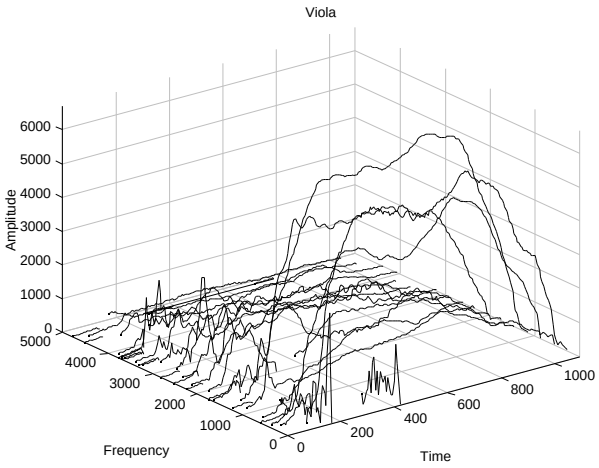


Figure 4.1. FFT analyzed additive parameters for the viola.

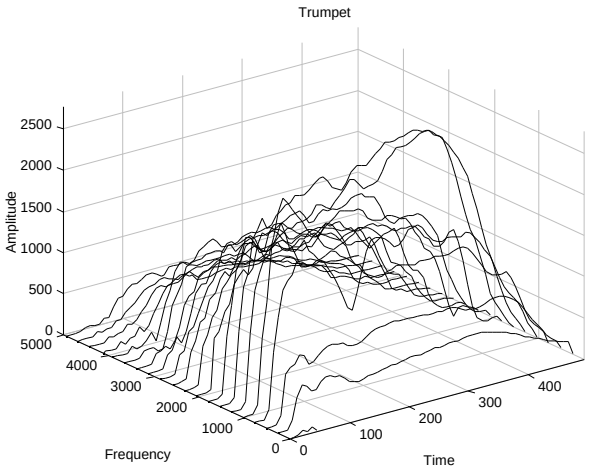


Figure 4.2. FFT analyzed additive parameters for the trumpet.

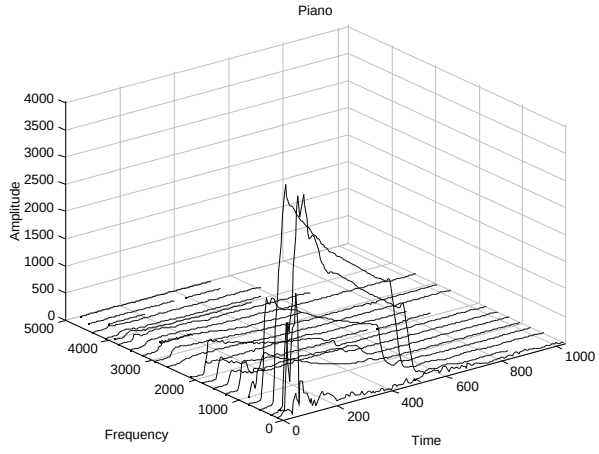


Figure 4.3. FFT analyzed additive parameters for the piano.

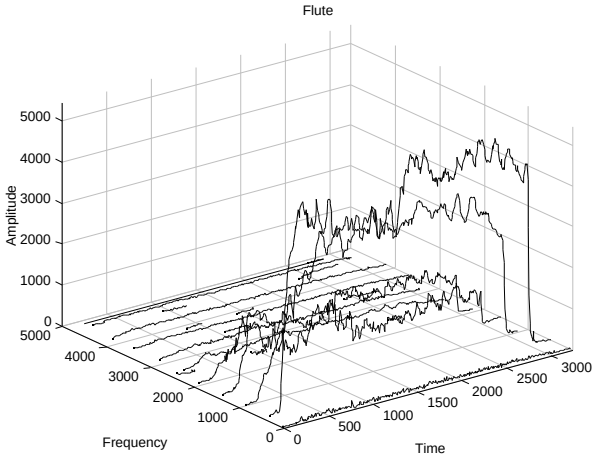


Figure 4.4. FFT analyzed additive parameters for the flute.

Although the result is generally satisfactory, it is clear that some phenomena are not well analyzed with this method. The attack of the piano sound seems blurred, and the noise in the flute has disappeared. The trumpet has a good resynthesis with the FFT-based additive analysis, and the viola also seems acceptable. The partial tracking still has some problems with low amplitude partials that come and go, especially in the flute sound. These are easily removed, and not shown in the FFT plots.

The main problem seems to be the lack of realism and presence in the resynthesized sounds. It seems related to the lack of fast transients and noise in the resynthesized sounds.

Visually, the viola seems to have a lot of transient behavior in the attack. Most sounds have a false partial in the high energy, low frequency region. The trumpet looks very nice, which also corresponds to the good quality of the resynthesis. The piano also looks very good, but the resynthesis is slightly dull and blurred.

The problems are caused by the limited time resolution in the FFT analysis, which is caused by the time/frequency limitation. This limitation states that the frequency support of a frequency domain window is inversely proportional to the time support of the inverse transform of the same window.

Therefore, to get a frequency domain window small enough, so that it does not touch the adjoining partials, a large time domain window is necessary, as can be seen in figure 4.5. The top plot shows 3 time domain windows, and the bottom plot shows the FFT of the same windows multiplied by a sinusoid. The time domain window should be small so that it discriminates between two close phenomena, but the frequency domain window also needs to be small, so it separates two close partials. This is mutually exclusive.

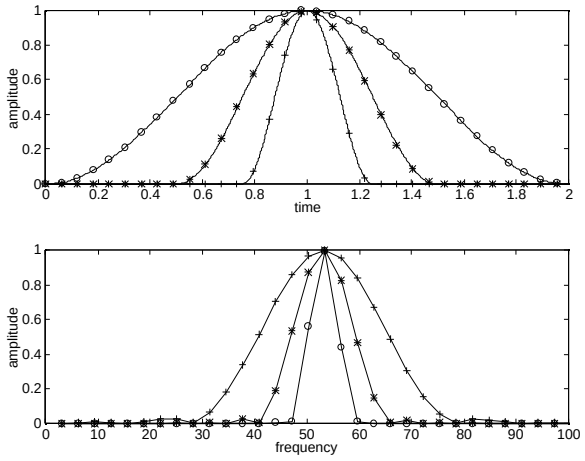


Figure 4.5. Illustration of the time / frequency window discrimination. A small time domain window yields a large frequency domain window, and vice verse.

4.3. Linear Time/Frequency Analysis

Philippe Guillemain [Guillemain 1994] has developed a solution to the time window/frequency window limitation. Here, the influence of adjoining partials is eliminated by putting loose limitations on the frequency domain behavior of a filter.

In figure 4.6 it is shown that even large frequency domain windows with a short time support, can be combined to the characteristics wanted. A linear combination of the dotted windows is used to create the solid line frequency domain filter.

Limitations are used to create the resulting filter. The limitations are that the filter are one at the frequency analyzed, and zero everywhere on all other frequencies.

It can be shown that a linear combination of gaussians conserves the same time-support as each gaussian individually.

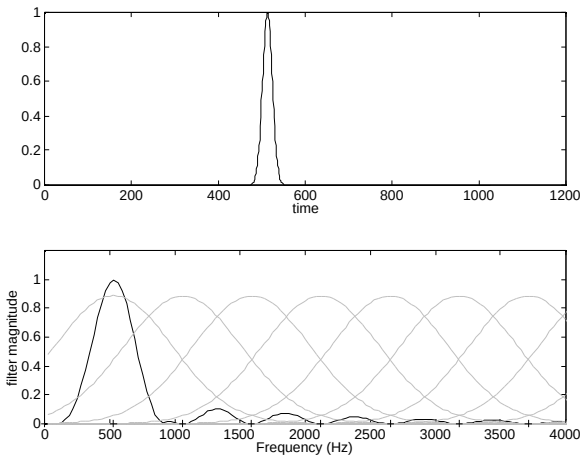


Figure 4.6. Principle of the LTF filter construction. Time domain (top) and frequency domain (bottom).

4.3.1. Constructing the Filters

The full theory of this analysis method can be found in [Guillemain 1994] and [Guillemain *et al.* 1996]. In the following, the zero-order filter is exposed. Zero-order meaning that the filter is one at the frequency being analyzed, and zero on all other initial frequencies. In the first order filter, the derivative is zero on all frequencies being analyzed. This stabilizes the filters, since it ensures that the frequency behavior is slowly varying in the neighborhood of the analyzed frequency.

The model of the signal $s(t)$ to analyze is a sum of sinusoidals,

$$s(t) = \sum a_k(t)\sin(\omega_k t) \qquad \textbf{(4.5)}$$

The variations of $a_k(t)$ are supposed to be much slower than $s(t)$.

The Gabor transform is performed on the signal,

$$L_g(\tau,\alpha) = \int s(t)W(t-\tau)e^{-i\alpha(t-\tau)}dt \qquad \textbf{(4.6)}$$

which can be found in the frequency domain, by the Parseval relation,

$$L_g(\tau,\alpha) = \int s(\omega)\hat{W}(\omega-\alpha)e^{i\omega\tau}d\omega \qquad \textbf{(4.7)}$$

since the signal is known, the discrete version of (4.7) is further developed,

$$L_g(\tau,\alpha) = \sum_{k=1}^{N} \frac{a_k}{2}\left(\hat{W}(\omega_k-\alpha)e^{i\omega_k\tau} + \hat{W}(-\omega_k-\alpha)e^{-i\omega_k\tau}\right) \qquad \textbf{(4.8)}$$

$$L_g(\tau,\alpha) = \sum_{k=1}^{N} \frac{a_k}{2}(\hat{W}(\omega_k - \alpha) + \hat{W}(-\omega_k - \alpha))\cos(\omega_k \tau) + i\sum_{k=1}^{N} \frac{a_k}{2}(\hat{W}(\omega_k - \alpha) - \hat{W}(-\omega_k - \alpha))\sin(\omega_k \tau) \quad \textbf{(4.9)}$$

Since $L_g(\tau,\alpha)$ is known, the equation (4.9) is further developed at the initial frequencies,

$$\sum_{k=1}^{N} \frac{a_k}{2}(\hat{W}(\omega_k - \alpha_p) + \hat{W}(-\omega_k - \alpha_p))\cos(\omega_k \tau) = \mathrm{Re}(L_g(\tau, \alpha_p)) \quad \textbf{(4.10)}$$

$$\sum_{k=1}^{N} \frac{a_k}{2}(\hat{W}(\omega_k - \alpha_p) - \hat{W}(-\omega_k - \alpha_p))\sin(\omega_k \tau) = \mathrm{Im}(L_g(\tau, \alpha_p)) \quad \textbf{(4.11)}$$

$1<p<N$. This gives two linear systems, N equations and N unknown variables, where N is the number of partials.

$$\sum_{k=1}^{N} W_{p,k} X_k(\tau) = L_g(\tau) \quad \textbf{(4.12)}$$

The elements of the first system is given by,

$${}^1W_{p,k} = \frac{1}{2}(\hat{W}(\omega_k - \alpha_p) + \hat{W}(-\omega_k - \alpha_p)) \quad \textbf{(4.13)}$$

and the elements of the second system is,

$${}^2W_{p,k} = \frac{1}{2}(\hat{W}(\omega_k - \alpha_p) - \hat{W}(-\omega_k - \alpha_p)) \quad \textbf{(4.14)}$$

The signal at frequency k can now be calculated,

$$X_k(\tau) = {}^1W_{p,k}^{-1}\,\mathrm{Re}(L_g(\tau)) + {}^2W_{p,k}^{-1}\,\mathrm{Im}(L_g(\tau)) \quad \textbf{(4.15)}$$

Remember that the signal is supposed to be a sum of sinusoidals, so

$$X_k(\tau) = a_k(\tau)e^{i\omega_k \tau} \quad \textbf{(4.16)}$$

The output vector k is thus a complex partial with amplitude a_k. The zero order analysis assumes that all time derivatives of a_k are zero. As a result, the analysis performs better when a_k is a smooth function.

In practice, the frequencies and the amplitudes are often extracted using a time domain or a frequency domain filter. Therefore (4.15) is developed. The resulting filters then become [Guillemain *et al.* 1996],

$$X_k(t) = \int \sum_{p=1}^{N} \begin{Bmatrix} {}^1W_{k,p}^{-1} \dfrac{\overline{\hat{W}}(\omega - \alpha_p) + \hat{W}(\omega + \alpha_p)}{2} \\ {}^2W_{k,p}^{-1} \dfrac{\overline{\hat{W}}(\omega - \alpha_p) - \hat{W}(\omega + \alpha_p)}{2} \end{Bmatrix} s(\omega) e^{i\omega\tau} d\omega \qquad \textbf{(4.17)}$$

$$X_k(t) = \int \sum_{p=1}^{N} F_k(\omega) s(\omega) e^{i\omega\tau} d\omega \qquad \textbf{(4.18)}$$

$F(\omega)$ is a filter banc of dimension N, with the following properties.

1) $|F_k(\alpha_p)| = \delta_{p,k}$, $|F_k(-\alpha_p)| = 0$, p, k [1, N]

2) The time support of F_k is equal to the time support of $W(t)$

3) $F_k(t)$ convoluted with $\sum_{k=1}^{N} A_k \sin(\omega_k t + \Phi_k)$ is $A_k e^{i(\omega_k t + \Phi_k)}$, k [1,N]

This signifies that the output of filter F_k is zero for all initial frequencies except f_k and that the time-support of F_k is equal to the time support of the window *W(t)*. Since no hypothesis has been made on the window, the frequency domain window *W(t)* can spread over several partials without ruining the good properties of F_k.

A gauss window is used in this work. The standard deviation of *W(t)* is found by setting the value of *W(t)* to a constant value at the closest neighboring initial frequency.

The time domain filter is used in this work, and it is,

$$F_k(t) = \sum_{p=1}^{N} W(t) \left({}^1W_{k,p}^{-1} \cos(\alpha_p t) + i\, {}^2W_{k,p}^{-1} \sin(\alpha_p t)\right) \qquad \textbf{(4.19)}$$

The frequency domain filter characteristics can be seen in figure 4.7 along with the log of the absolute spectrum of a flute sound. This filter has the characteristics wanted, it is 1 at the frequency analyzed and 0 for all other frequencies. The filters have rebounds between the initial frequencies, but this does not generally disturb the analysis. This zero-order filter does not guarantee smooth response at the initial frequencies.

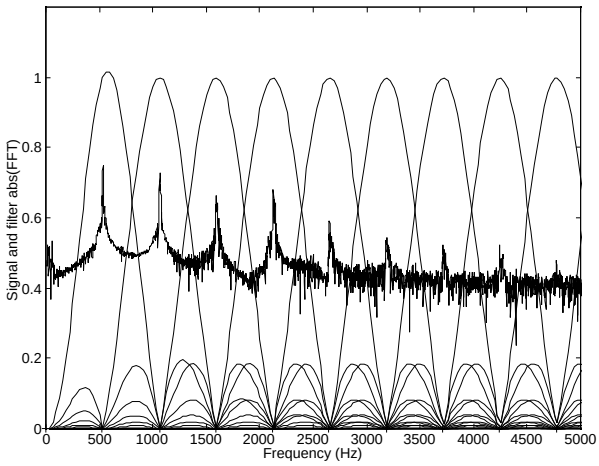


Figure 4.7. Zero-order filters and signal FFT for a flute sound.

The filters can be improved by setting the frequency derivatives of the filters to zero at the initial frequencies. The additional conditions are used to model amplitudes with nonzero derivatives.

The filters used in the rest of this work are the filters of order 1. See [Guillemain 1994] and [Guillemain *et al.* 1996] for details on higher order analysis.

Before the signal is analyzed with the filter, the initial frequencies, that is the frequencies being analyzed, must be determined.

4.3.2. Initial Frequencies

In order to perform the analysis, it is necessary to know the frequencies that are interesting in the sound. They are found by the initial frequency estimation presented in the fundamental frequency estimation in Chapter 3. The initial frequencies now consist of all the quasi-harmonic frequencies and strong non-harmonic frequencies.

The estimated frequencies are used in the time-frequency analysis, since they are believed to be better than the FFT-analyzed frequencies which are sometimes misjudged, and which are often missing some harmonic components.

With the introduction of initial frequencies, the filters can be constructed, as described in paragraph 4.3.1. The frequencies and amplitudes of the partials are extracted in paragraph 4.3.4, but first a method for the elimination of rebounds of the filters is presented.

4.3.3. Rebounds

Although the filters used in the analysis have the desired properties, notably, the frequency response is one at the frequency being analyzed, and zero at all other frequencies, they sometimes introduce frequency domain rebounds, which, if they are positioned in strong noise, can ruin the analysis. In figure 4.8 the rebounds can clearly be seen in the noisy low-frequency range of the piano sound.

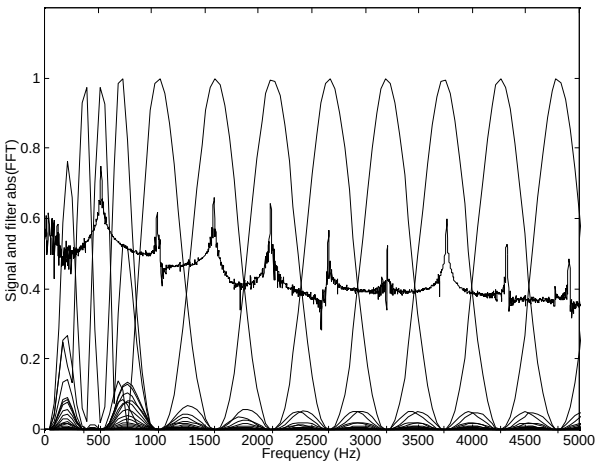


Figure 4.8. Rebounds of filter.

Since the noise is much stronger than the weak upper partials, it can sometimes disturb the analysis. The frequencies of some of the partials of the piano can be misjudged due to the hammer noise present in the rebound frequency area. By misjudged is meant that the analyzed frequency is no longer close to the initial frequency. Although this may be desirable in some situations, especially in a direct resynthesis, where the mechanical noise is restituted with the harmonic components, it makes some simple operations on the partial frequency impossible, such as the estimation of the mean frequency.

Therefore, a method of eliminating these rebounds is introduced. It eliminates the rebounds in strong noise so the estimation of the partial amplitude and frequency is not influenced by the noise component.

In order to minimize the disturbance of the elimination, it is done only when necessary, i.e. when the rebound amplitude is much larger than the partial amplitude. The filter and the signal are transformed into the frequency domain by FFT, and multiplied. To have the same number of points, the maximums of every N points of the signal FFT are used,

$$sf = FFT(filter)\ \ FFT(signal) \qquad \textbf{(4.20)}$$

The maximum of $|sf|$ outside the partial being analyzed is now compared with the maximum of $|sf|$ in the frequency being analyzed. If it is relative strong, the filter at the strong frequency is multiplied with an inverted hamming the size of the fundamental. sf is calculated again and the process is repeated until there are no more strong amplitudes outside the frequency being analyzed.

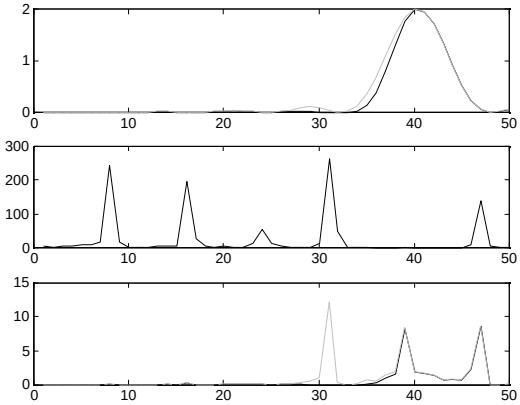


Figure 4.9. Detail of the FFT of Filter (top), signal (middle) and result of filtering for the fifth partial of the piano sound. x axis is frequency bins, and y-axis is amplitude. Original (dotted) and after elimination of rebounds (solid).

The original and manipulated filters can be seen in figure 4.9 (top). The FFT of the signal is shown in figure 4.9 (middle). It is clear that the fifth partial is very weak.

The magnitude of the fifth partial analyzed with the original filter, and with the manipulated filter (dotted), can be seen in figure 4.9 (bottom). The influence of the strong fourth partial has been eliminated.

The first 5 partial frequencies of a piano sound, as analyzed by the filter without rebound elimination can be seen in figure 4.10 and the same 5 partial frequencies as analyzed by the filter with rebound elimination can be seen in figure 4.11.

The elimination of rebounds has clearly succeeded. The dip in the middle of the first half of the fifth partial has disappeared, and there seems to be less noise in the silent second half.

The elimination of rebounds should not be done in an analysis/synthesis situation, but only when other features are to be extracted from the partials, such as the mean frequency. In such a case, the elimination of the rebounds stabilizes the feature extraction.

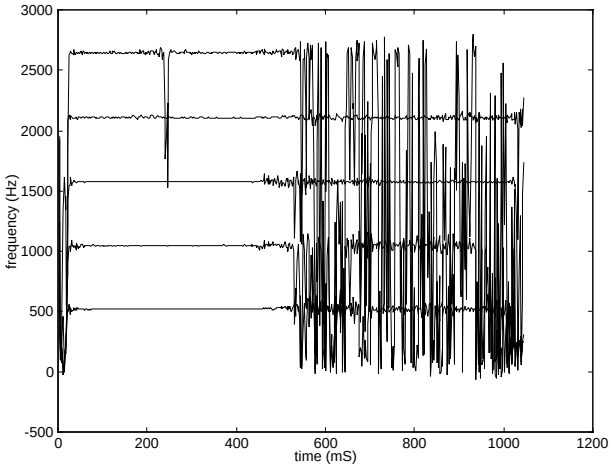


Figure 4.10. The 5 first harmonic overtones of piano C4 sound.

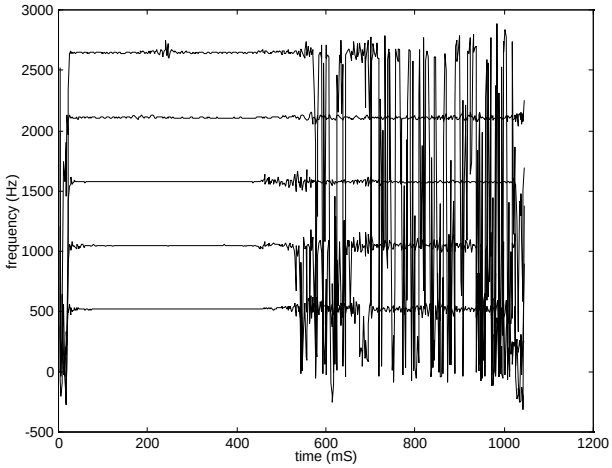


Figure 4.11. The 5 first harmonic overtones of piano C4 after elimination of rebounds.

4.3.4. Frequency and Amplitude Extraction

The filters from paragraph 4.3.1 can now be constructed with the initial frequencies found in paragraph 4.3.2. The output of filter *k* is,

$$X_k(t) = s * F_k = \int s(\tau)F_k(t-\tau)d\tau \qquad \textbf{(4.21)}$$

and the corresponding frequency and amplitude are

$$f_k(t) = \frac{s_r}{2\pi}\frac{\partial}{\partial t}\left(\arctan\left(\frac{(X_k(t))}{(X_k(t))}\right)\right) \qquad \textbf{(4.22)}$$

$$a_k(t) = |X_k(t)| \qquad \textbf{(4.23)}$$

The phase must be unwrapped before the frequency difference is calculated.

4.3.5. Data Reduction

The resulting partial frequencies and amplitudes can be obtained for each sample, but this resolution introduces too much data (more than the sampled sound), and a data-reduction scheme is necessary. Two approaches have been tested, piecewise linear approximation [Bernstein *et al.* 1976], [Horner *et al.* 1996] and averaging over one period [Grey *et al.* 1977]. Although piecewise linear approximation potentially has a better data reduction, the method seems to introduce artifacts in the sound when modeling noise in the additive parameters, and the simpler averaging over one period was chosen. This data reduction method is also justified by the time resolution of this analysis, as shown in section 4.4. The period averaging, which guarantees synchronous partials, also simplifies subsequent operations on the additive parameters [Wessel 1997].

4.4. Comparison of FFT and LTF Analysis

The LTF analysis seems to render a more faithful reproduction of the sounds. This is believed to be due to the better time-resolution of this analysis method. In order to compare the time resolution of the two analysis methods, a few test signals have been created and analyzed with the two methods. The calculated time-resolution and mean square of the error results are then compared.

4.4.1. Test Signals

The test signals created are 4 one-second sounds with 8 harmonic partials. All partials have the same amplitude slope, first 1/8-second silence, then a linear slope from maximum amplitude to 1/3 of maximum amplitude for 3/4 seconds, and then 1/8-second silence. The sounds have fundamental frequencies 30 Hz, 100 Hz, 300 Hz and 1000 Hz. For an example of the test signal, see figure 2.1.

4.4.2. Analysis

The test sounds are first analyzed using the FFT and the linear time/frequency (LTF) analysis. No smoothing has been done on the LTF analyzed parameters. The resulting spurious partials are removed. The rise time and the fall time of each partial, defined as the time between 10 % and 90 % of the amplitude, are calculated. Furthermore, the frequency and amplitude errors are defined as the mean of the square of the error between the original

partial and the analyzed partial and normalized by the maximum amplitude and the mean frequency respectively.

The amplitude and frequency errors are,

$$f_{error} = \frac{1}{N_{sustain}} \sqrt{\sum \left(\frac{f_{sustain} - mean(f_{sustain})}{mean(f_{sustain})} \right)^2} \quad \textbf{(4.24)}$$

$$a_{error} = \frac{1}{N_{sustain}} \sqrt{\sum \left(\frac{a_{sustain} - ramp(a_{\max}, a_{\max}/3)}{a_{\max}} \right)^2} \quad \textbf{(4.25)}$$

4.4.3. Results

The time resolutions for the attack (top) and the release (bottom) are showed in figure 4.12. The ‘o’ are the FFT times, and the ‘*’ the LTF times. The errors are shown in figure 4.13 for the amplitude (top) and the frequency (bottom). The plots are made for the mean of the error of the eight partials.

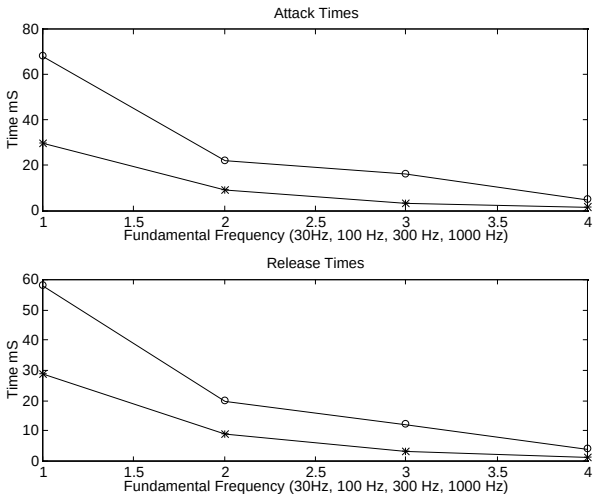


Figure 4.12. Time resolution for 4 test signals. FFT analysis is ‘o’ and LTF analysis is ‘*’.

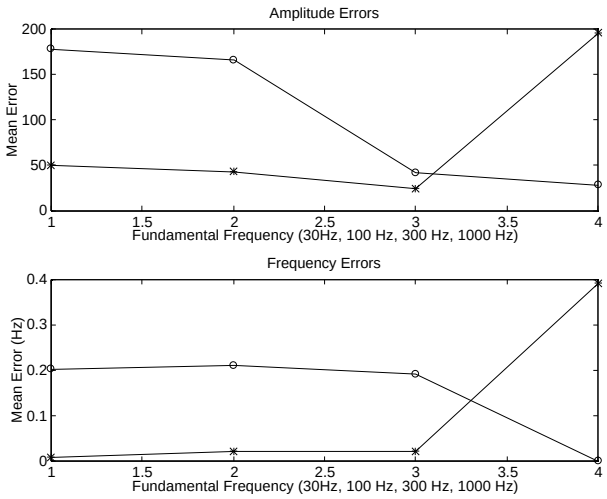


Figure 4.13. Amplitude error (top) and frequency error (bottom) for the FFT analysis ‘o’ and the LTF analysis ‘*’.

The time-resolution for the FFT is about twice the period time, which corresponds well with the fact that a hamming window of four times the period is used, and the effective time support is about half the window length. The LTF time resolution is about equal to the period length, which is about twice as good as the FFT time resolution.

The amplitude and frequency error is generally smaller for the LTF analysis method than for the FFT-based analysis. The frequency error is relatively constant over the frequency range. One exception is the 1 kHz test signal, where the LTF analysis performs badly, probably due to overshot of the short filters, and the FFT performs very well,

perhaps because the frequencies fall exactly on the frequency bins of the FFT. The LTF filters are of order 1 and the discontinuity at the slope edges are not well handled by this method. This is not a problem generally, since real life signals generally are smooth enough.

4.5. Resynthesis

The additive parameters for four sounds analyzed with the LTF method are shown in figure 4.14 to figure 4.17.

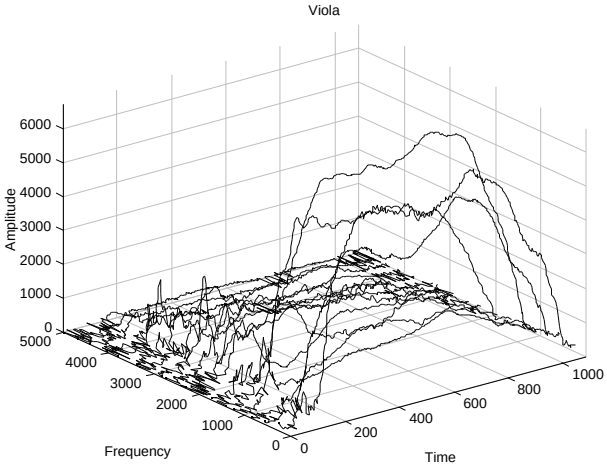


Figure 4.14. LTF based additive parameters for the viola.

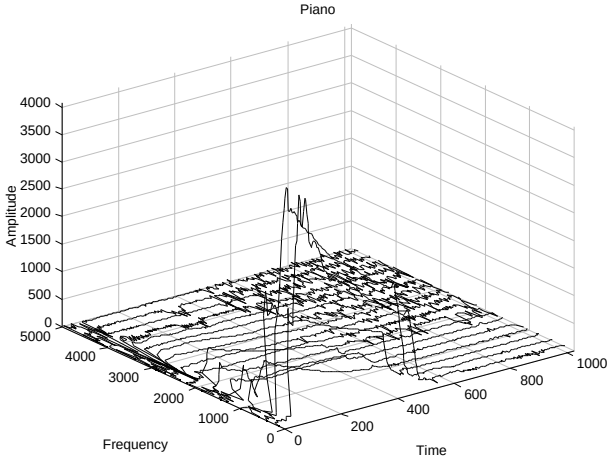


Figure 4.15. LTF based additive parameters for the piano.

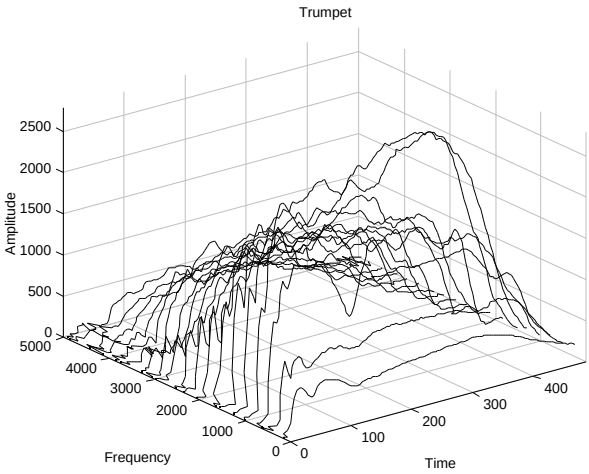


Figure 4.16. LTF based additive parameters for the trumpet.

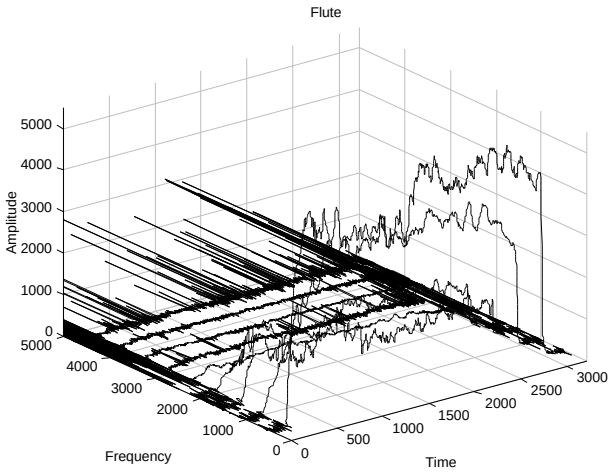


Figure 4.17. LTF based additive parameters for the flute.

Visually, the main difference between the LTF parameters and the FFT parameters in figure 4.1 to figure 4.4 are the noise on the high partials of the flute, and the faster irregularities in the viola and the trumpet. Furthermore, the LTF analysis does not have any spurious frequencies, or tracks that are born or die in the middle of the sound. The noise in

the LTF analysis is present in the additive parameters. The sum of the magnitude of the analyzing LTF filters is close to one for all frequencies, which guarantees the reconstruction of the additive noise.

The sounds of the LTF analysis are definitely better than the sounds from the FFT-based analysis. The main differences are the additive noise in the flute sound, and the more distinct attack of the piano. Generally, the LTF method seems to render a more realistic resynthesis. The sounds have more presence, and a greater realism, as compared with the sounds from the FFT-based analysis. Some of the degradations in the resynthesis can probably be attributed to the omission of phase information. [McAuley *et al.* 1986] has compared the resynthesis with and without phase information. Their conclusions are that the omission of phase information made the resynthesis different than the original, whereas the resynthesis with phase information was not. This was more pronounced for low-pitched voices. They also model noisy speech and found that “the noise took on a tonal quality that was unnatural and annoying”, if the phase information was not used. Phase has not been included in this work.

Since phase was used in none of the analyzing methods in this chapter, the conclusion is still that the LTF analysis performs significantly better than the FFT-based analysis.

The linear time-frequency analysis method seems well adapted for musical sounds. Because of the looser constraints in the frequency domain, it has a better timing resolution than the FFT. This timing resolution obviously better models fast transition, but it also permits the analysis of noise, both variations in the partial amplitude (shimmer) and variations in the partial frequency (jitter) [Richard *et al.* 1996]. Its good timing resolution permits a successful analysis of traditionally difficult phenomena, such as the fast attack of the piano, and the additive noise in the flute.

4.6. Conclusions

In this chapter, two methods for additive analysis of musical sounds are compared. The conclusion is that a new method, called linear time-frequency analysis (LTF) performs significantly better than the classical FFT-based analysis. Details were given for both methods, and this work presents a new method for the restitution of good frequency analysis of weak partials for the LTF analysis.

The analysis methods were compared, both the time resolution and the mean square amplitude and frequency errors were calculated for several test signals for both methods.

The LTF analysis has twice as good time resolution, with significantly lower frequency and amplitude errors.

The quality of the resynthesis of the LTF analysis is very good. Even though the partial amplitude and frequency are averaged over one period, it is virtually impossible to distinguish between the original and the resynthesized sound. However, some doubt can still be expressed as to whether the phase should be included in the additive parameters. More work remains before this issue has been resolved. It seems clear (cf. Chapter 2) that the ear is sensitive to differences in phase, at least for low frequencies.

The LTF analysis presented in this chapter is used in the following chapters for the analysis of musical sounds. The FFT-based method is not used in the rest of this work.

Chapter Five

5. Envelope Modeling

This chapter models the envelope of the partials. The envelope is the evolution over time of the amplitude of a sound. It is one of the important timbre attributes. A faithful reproduction of a noiseless sound with no glissando or vibrato can be created using the individual amplitude envelopes of the additive parameters. Unfortunately, the analyzed amplitude envelopes often contain too much information to be easily manipulated. A model of the envelope is therefore necessary.

The envelope model presented here is relatively simple, having only 4 split-points. The main characteristics of this model is the attack, the sustain or decay, and the release.

Two methods for the extraction of the attack and release times are compared: one method finds the envelope times by comparing the amplitude with a percentage of the maximum of the envelope, and a new method developed here finds the envelope times by analyzing the derivatives of the amplitude. This method performs significantly better than the classical percent-based method.

5.1. Introduction

The modeling of amplitude or other time-varying parameter in discrete time/value pairs is as old as electronic music. The ADSR envelope generator, which was introduced with the first analog synthesizers, divides the envelope in four steps, Attack, Decay, Sustain and Release, see figure 5.1. The ADSR approach relies on an exponential envelope, which corresponds well with the perceptual quality of the amplitude.

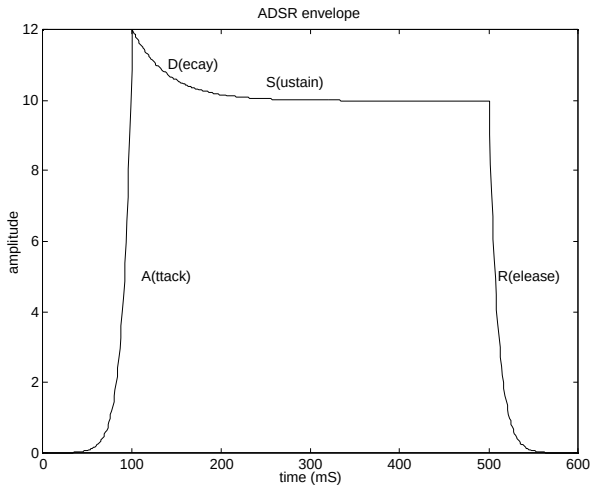


Figure 5.1. ADSR envelope.

Generally though, the ADSR model is used only on 'total' control parameters, such as amplitude, or filter frequency, and not on individual additive partials.

The additive parameters have traditionally been modeled in line segments [Bernstein *et al.* 1976]. The idea is that a continuous curve can be simplified in a series of line segments and the error, which is the difference between the continuous curve and the line segment curve, is negligible. See figure 5.2 for an illustration of the line-segment approximation. The crosses indicate the split points and the original envelope is shown in a dotted line. The number of line segments is a function of the maximum error tolerated. There are 8 line segments in figure 5.2.

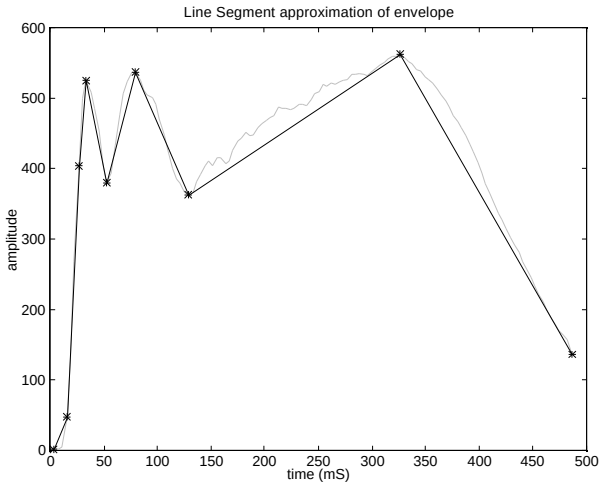


Figure 5.2. Line segment approximation of envelope.

The envelope of a musical sound has been the object of many studies. [Schaeffer 1966] proposes a classification of attack genres. [Freedman 1967] models the envelope as cascading exponentials. [Strong *et al.* 1967] models the envelope in linear segments multiplied with a smoothed error. [Tellman *et al.* 1995] models the envelope with an attack-decay model including other features, such as tremolo.

Analysis of the envelope parameters can be used in auditory perception research. [Gordon 1987] analyses the perceptual attack time of musical tones. [Krimphoff *et al.* 1994] correlates measured attack times with perceptual input from listening tests.

Physical models of musical instruments show that the decay of a flute tube excited by a Dirac is an exponential [Ystad *et al.* 1996]. This is also the case for the guitar (and other string instruments) [Karjalainen *et al.* 1993].

The envelope model can be seen as a data reduction of the additive parameters. [Strawn 1980], [Charbonneau 1981] and [Horner *et al.* 1996] compare different envelope approximations.

The model introduced in this work combines the intuitive simplicity of the ADSR model with the flexibility of the additive model. The idea is to model each partial amplitude as four time/value pairs, here called start of attack (soa), end of attack (eoa), start of release (sor) and end of release (eor). Furthermore, the interval between each split point is modeled by a curve the quality of which (exponential/logarithmic) can be varied with one parameter. The soa, eoa, sor, eor model corresponds to the physical act of introducing energy into a system for a certain time. The difference between the start and the end of attack (attack time) and release (release time) is thus the time it takes the system to settle for this partial. This is the attack-sustain-release type of sound. If instead energy is introduced only once, such as in the plucked string, and the system is later damped, the attack-decay-release type of sound is produced. The model presented here models both types of sounds.

This model does not take into account tremolo or other effects. The sounds are supposed to be glissando-, vibrato- and tremolo-free, but these effects can be added to the additive parameters at any time.

This chapter first describes the timing extraction in section 5.2 with two different methods, and a comparison between the methods. The curve form between the split points is modeled in section 5.3. The envelope reconstruction is presented in section 5.4 and the additive parameters are created from the envelope parameters in section 5.5. Some novel ideas on the sharpening of the envelope are presented in section 5.6 and the chapter finishes with a conclusion.

5.2. Timing Extraction

The envelope times are important timbre attributes. The attack time, for instance, is recognized as one of the most important timbre attributes [Krumhansl 1989], [McAdams *et al.* 1995]. The envelope times found here are the start and end of the attack and release.

Two methods have been tested for the extraction of envelope times. The first, here dubbed the percent method, consists of finding the maximum of the curve to be modeled, and then finding the first or last time in the curve where the value is above a constant percent of the maximum. This method has several drawbacks: it is sensitive to noise, and it doesn't really model the release time consistently if the sound is of the decay type. For these reasons, a new method has been developed, called the slope method. Here the attack and release are found by searching for the maximum and minimum of the derivative of the curve. The start and end of the attack and release are found by following the derivative until it is less (more) than a constant value times the maximum (minimum). To reduce noise sensitivity, the slope method is performed on a smoothed curve, and the times are followed through the less and less smoothed curve until the unsmoothed case.

5.2.1. Percent Method

The percent method consists of finding the maximum of the curve and then finding the times where the amplitude is higher than a certain percent. The percent method has been used in [Krimphoff *et al.* 1994] to correlate the perceptive dimension attack with the measured values.

The percents chosen here are 10% for the start of attack (soa) and the end of release (eor) times, 90% for the end of attack (eoa), and 70% for the start of release (sor).

Given the amplitude of one partial, the soa time is the first time the amplitude is above 10%, the eoa time is the first time the amplitude is above 90%, the sor is the last time the amplitude is above 70% and the eor time is the last time the amplitude is above 10%.

As can be seen in figure 5.4, the amplitude evolution is quite different for the four instruments being analyzed. The piano has a relatively fast attack, and a typical decay-release form, the release occurring when the damper is placed on the strings. It is very easy to see the release time at circa 500 mS.

Unfortunately, the release times for the piano analyzed with the percent method shown in figure 5.3 doesn't correspond very well the times observed in figure 5.4. It is not the

same for all partials, varying from almost 200 mS for the fundamental to less than 100 mS for the highest partials in a non-continuous manner.

The times found by the percent method are plotted in figure 5.3 for the viola, the trumpet, the piano and the flute. The x-axis is the partial index, and the y-axis is the time. The different curves are the split-point times, the lowest being the soa, followed by the eoa, the sor, and the highest one, the eor. Only quasi-harmonic partial times are shown.

The durations of the sounds are easy to see, it is about 1 second for the viola, 500 mS for the piano and the trumpet and 3 seconds for the flute.

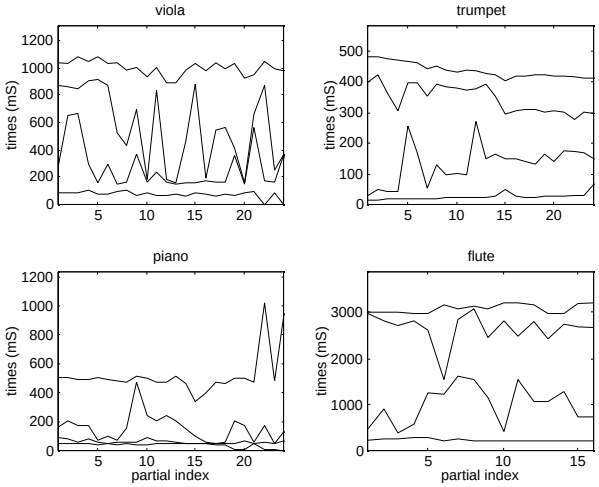


Figure 5.3. Percent times for the viola, the trumpet, the piano and the flute.

Although it is possible to adjust the percents so as to find the correct sor time for one partial, it is impossible to find the good sor time for all partials with the percent method, due to the difference in slope in the decay part.

The trumpet sound has the typical trumpet evolution, with different attack and release times for the different partials, the low partials starting faster and ending later than the high partials. In fact, there is no constant soa time as for the flute and the piano, but more like a continuous slope from circa 10 mS for the fundamental to almost 50 mS for the highest partial.

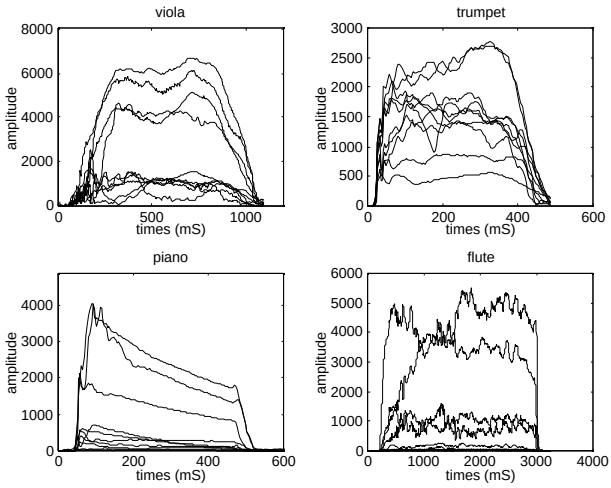


Figure 5.4. Amplitude curves for the viola, the trumpet, the piano and the flute.

Although the viola, trumpet and flute times are better than the piano times, there is a lot of noise on the times.

In conclusion, the percent times seem to correspond rather badly with what is observed in figure 5.4; they are noisy and the piano release times are completely wrong.

5.2.2. Slope Method

In the slope method, the envelope times are found by analyzing the derivative of the amplitude of the partial analyzed. The attack is found where the derivative is maximum in the first half of the sound, and the release is found where the derivative is minimum in the last half of the sound. In the initial search for times, done on a heavily smoothed envelope, the extremes of the attack and the release are found when the derivative is a constant multiplied by the maximum of the derivative. There are two constants, one for the soa, eoa and the eor, and one for the sor. The first constant, which is close to zero, model a curve which either starts or ends in zero, or ends in maximum, just after the attack. Therefore, the derivative is here zero-positive, whereas the sor curve can sometimes be, as in the piano release, a slow slope to a fast slope, in which case the derivative goes from one negative value, which indicates sustain, to a larger negative value which indicates release. Therefore the second constant is larger than the first constant.

The slope method is then; first find the maximum of the derivative, which corresponds to the middle of the attack, at_m

$$at_m = \max\left(\frac{\partial}{\partial t} envelope_{smoothed}\right) \quad \textbf{(5.1)}$$

then follow the derivative both backward and forward in time until it is smaller than a constant multiplied with the maximum of the amplitude. This is the start and end of the attack. The same is done for the release, although it is here the minimum value of the derivative that is searched, and the middle of the release, rt_m, that is found,

$$rt_m = \min\left(\frac{\partial}{\partial t} envelope_{smoothed}\right) \quad \textbf{(5.2)}$$

The principle is illustrated in figure 5.5, where the smoothed envelope (top) and the first derivative (bottom) are shown for the viola, the piano, the trumpet and the flute. The start and end of the attack and release are indicated with ‘+’. It can clearly be seen that the soa/eoa/eor times are much closer to the zero of the derivative, and thus closer to the end of the slope, whereas the sor time has a larger negative derivative, which permits both the analysis of attack-decay-release and attack-sustain-release type of sound.

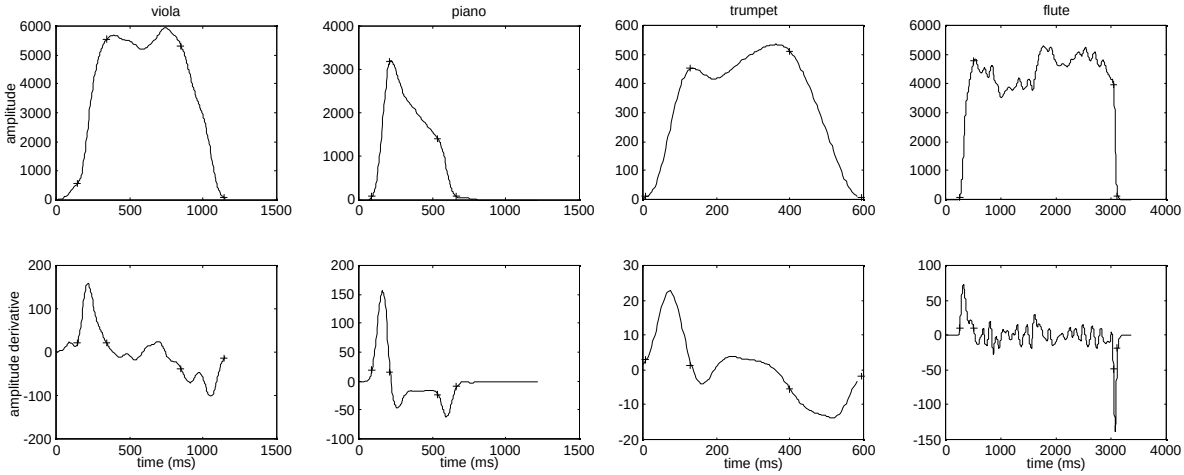


Figure 5.5. Amplitude and first derivative for the smoothed fundamental of four sounds with envelope times found with the slope method.

Notably, this envelope analysis method captures the decay release split point of the piano perfectly. All times seem to be close to the edges of the attack and the release and the higher constant used in the search for the start of release don't seem to disturb the non-decay sounds very much. Of course, the times found in the heavily smoothed case don't correspond to the times in the unsmoothed case, and it is thus necessary to 'follow' the times from the smoothed to the unsmoothed envelope.

The smoothing is done by multiplying the FFT of the envelope by the FFT of a gaussian. The larger the gaussian, the more smoothed the envelope. The method for following the points from the smoothed to the unsmoothed envelope has been borrowed from the scale-space theory used in image processing [Lindeberg 1996].

Scale-space is a model of the blur of the images seen at different distances. The blur is modeled by convoluting the images with a gauss with a variable standard deviation. Large structures can then be found if the standard deviation is large (images seen at a distance) and details can be found if the standard deviation is small (images seen at close range).

Many methods developed in the scale-space community could potentially find use in music informatics research, including the edge-detection following used here, but also top-point classification [Johansen 1994], deblurring and anisotropic filtering. Deblurring is used in section 5.6 and anisotropic filtering has been tested in Chapter 3.

The method for following the envelope times from the smoothed to the unsmoothed case is summarized below.

The local maximums and minimums of the second derivative of the envelope are found, typically by searching the zero-crossing of the third derivative. This corresponds to the end

points of a slope, as can be seen in figure 5.6. When the envelope is less smoothed, the slope is steeper, and the slope points found correspond more to the unsmoothed case. In the unsmoothed case, there are typically many points, as can be seen in figure 5.7. It is thus necessary to use enough smoothing steps so the slope points can be followed.

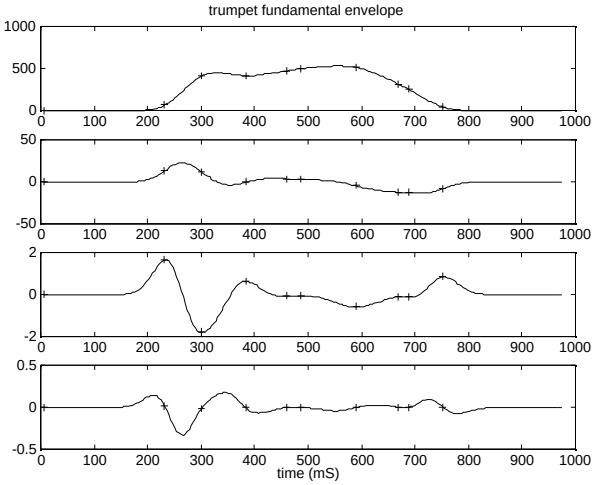


Figure 5.6. Envelope of smoothed trumpet fundamental and first three derivatives with the slope points.

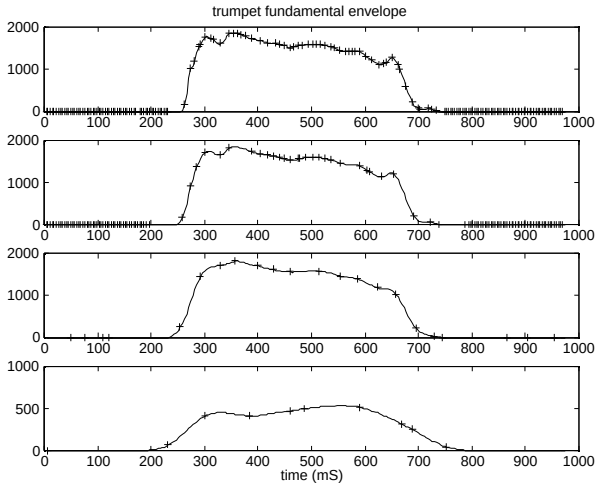


Figure 5.7. Slope points in different smoothing of the trumpet fundamental. Unsmoothed (top) to very smoothed (bottom)

Nevertheless, the times are adjusted after each smoothing step so it doesn't occur in the middle of the slope, or in a local minimum. Furthermore, if the slope point is chosen from many candidates, the closest to the middle of the attack (or release) is selected. This ensures that the attack and release get shorter in the unsmoothed case, as they should.

The resulting times for the viola, the trumpet, the piano and the flute are shown in figure 5.8.

The main difference between these times and the times found by the percent method in figure 5.3 is the start of release (sor) times found for the piano. As can be seen, the sor times now correspond roughly to the times that can be seen in figure 5.4. Furthermore, all times seem less noisy, and the behavior of the times is clearer now.

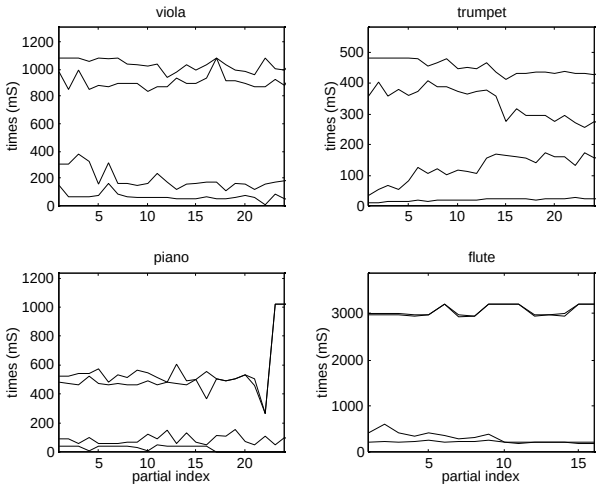


Figure 5.8. Slope times for the viola, the trumpet, the piano and the flute.

For instance, it is clear now that the flute attack times are shorter for the high partials than for the low partials.

The envelope time errors caused by noise in the weak upper partials are not present in figure 5.8.

5.2.3. Percent vs. Slope

Generally, the times from the slope method seem less noisy than the times found with the percent method.

It is also obvious when studying in detail the split points from the percent and the slope method, that the slope method finds split points much closer to the edge of the envelope, whereas the percent method never falls exactly on the edge. Therefore, the curves between the split points in the slope method become closer to the natural envelope, whereas the envelope of the percent method sometimes also contains the ends of the adjoining envelope slopes. The percent method is also sensitive to higher peaks, or noise, in the middle of the envelope. This can cause the percent method to indicate that both the attack and the release are positioned close to such a peak. In conclusion, the slope method seems more accurate and less noise sensitive than the percent method.

5.2.4. Relative Amplitude (percents)

In the slope analysis, the envelopes have variable amplitudes at the split points as opposed to the percent method, where the amplitude is a fixed percent of the maximum of the amplitude of the partial. The variable sor percents permit the modeling of both sustained or decaying sounds.

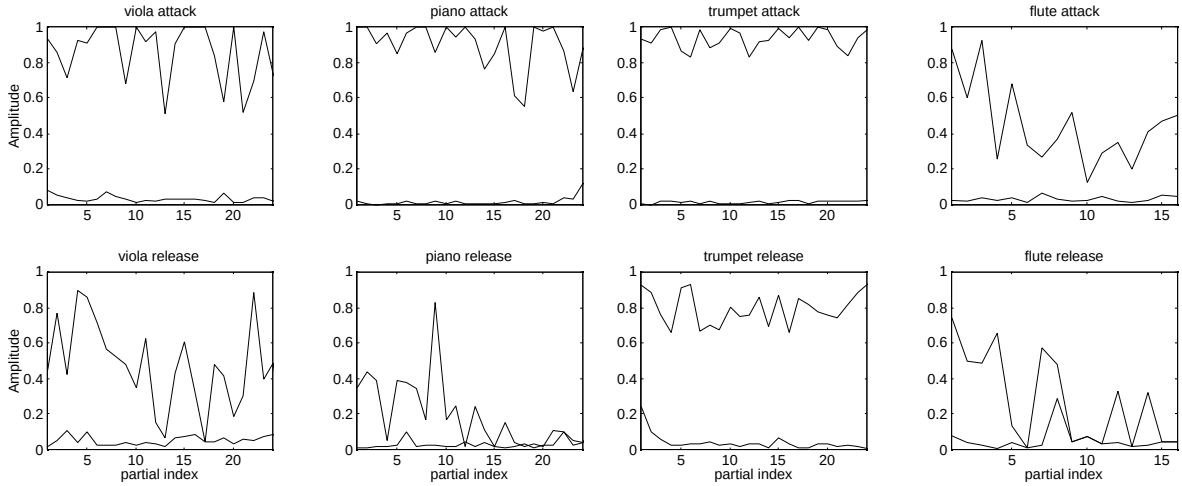


Figure 5.9. Attack (top) and release (bottom) percents for the four instruments.

The relative amplitudes, that is, the split point amplitudes divided by the maximum of the envelope amplitudes are shown in figure 5.9 for the attack (top) and release (bottom).

The x-axis is the partial index and the y-axis is the relative amplitude (percents) at the split points. The two lines in each plot are the start and end percents of each curve (attack or release).

The trumpet values seem very good, as does the piano attack. The piano release is much lower, since it is the time of release of the note that has been found where the strings already have lost some energy in the decay segment. This seems true for all instruments to a lesser degree, although it is probably caused by the analysis derivative threshold, which is higher for the release.

5.3. Curve Form

An estimation of the envelope times is now available, but the curve between the envelope points is not known. The evolution between the envelope points is modeled by a curve which has parameter defined exponential/logarithmic slope. This curve presumably models all the curves possible. Obviously, no oscillation or irregularity is modeled, but these are assumed to be either tremolo or noise. Tremolo is not modeled in this work and the noise model is presented in Chapter 6.

There are five segments with a curve form for each partial; the start, attack, sustain, release and end segments.

5.3.1. Curve Model

No hypothesis is made on the slope of the envelope curve between split points. Instead, the slope is modeled by a curve whose curve form can be set with one parameter.

The curve used for the modeling of the envelope for one segment is

$$Curve_s = v_0 + (v_1 - v_0)(1 - (1 - x)^n)^{\frac{1}{n}} \qquad \textbf{(5.3)}$$

The *x* value is normalized between zero and one. The value of *n* is always positive. Another curve form, which may have a more physical relevance, is the exponential curve,

$$eCurve_s = v_0 + (v_1 - v_0)\frac{e^{n\,x} - 1}{e^n - 1} \qquad \textbf{(5.4)}$$

Unfortunately, no resynthesis comparisons have been made between the two curves. The two curves are quite similar, but the equation (5.4) has the problem of being undefined when *n* equals zero.

The exponential curve is probably preferable from a physical point of view, but the curve from equation (5.3) has been used in the rest of this work.

The form of this curve can be seen in figure 5.10. The slope form changes as a function of n. When *n* is close to zero, the curve is exponential, when *n* is one, the curve is linear, and the curve is logarithmic when *n* is greater than one. This curve should now be fitted to the envelopes between the envelope times found.

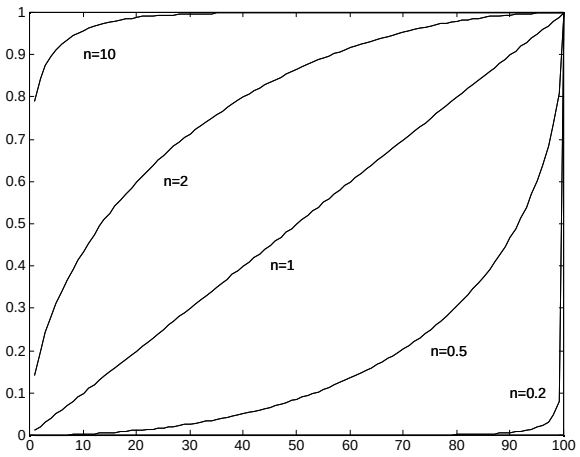


Figure 5.10. Different slopes for the envelope curve going from 0 to 1.

5.3.2. Language Conventions

The curve forms of the envelopes of many sounds are analyzed in the following chapters, and in order to understand the analysis, the appellations for the different curve forms must be clear.

The different curve forms possible are shown in figure 5.11 for the attack (top) and release (bottom).

The attack, or any positive slope, is said to be logarithmic when $n>1$ and exponential when $n<1$. The release, or any negative slope, is said to be exponential when $n>1$ and logarithmic when $n<1$.

Furthermore, an attack with curve form value n_1 is said to be more exponential than another attack with curve form value n_2 if $n_1<n_2$, or more logarithmic if $n_1>n_2$.

A release with curve form value n_1 is said to be more exponential than another release with curve form value n_2 if $n_1>n_2$, or more logarithmic if $n_1<n_2$.

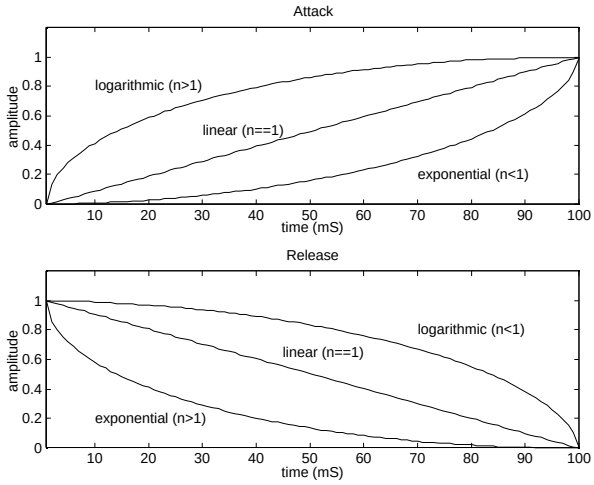


Figure 5.11. Possible curve forms for the attack(top) and release (bottom).

5.3.3. Curve Fitting

The curve form value *n* is found by minimizing the least-square error,

$$Error = \sum_{t=1}^{N} \left(Curve_t - Envelope_t \right)^2 \qquad \textbf{(5.5)}$$

The curve-fitting problem is nonlinear, and the Levenberg-Marquardt method is used to solve it.

Implementation details can be found in [Moré 1977] and an example of the curve found can be seen in figure 5.12. The curve value *n* is 1.8.

The sustain segments that have the same amplitude at the start and end have no defined curve form values, which are then random and often quite large. This is often the case for the start or end segments as well.

The attack and release segments curve form values are generally well defined.

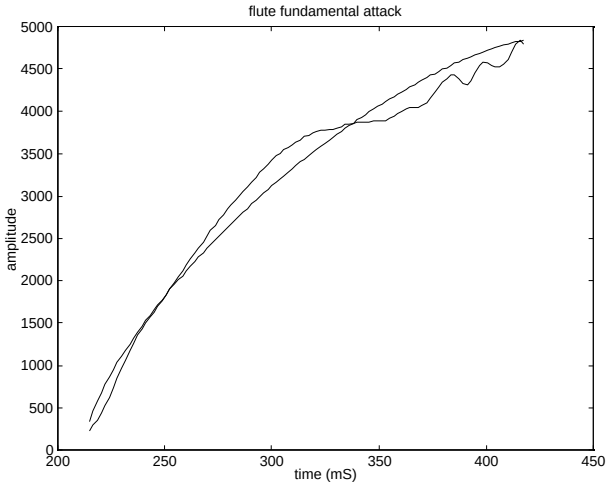


Figure 5.12. Curve Fitting for the attack of the trumpet fundamental.

The curve form values for the attack, sustain and release for the viola, trumpet, piano and flute are plotted as a function of the partial index in figure 5.13. The top plots are the attack curve form values, the middle plots are the values for the sustain, and the bottom plots are the values for the release.

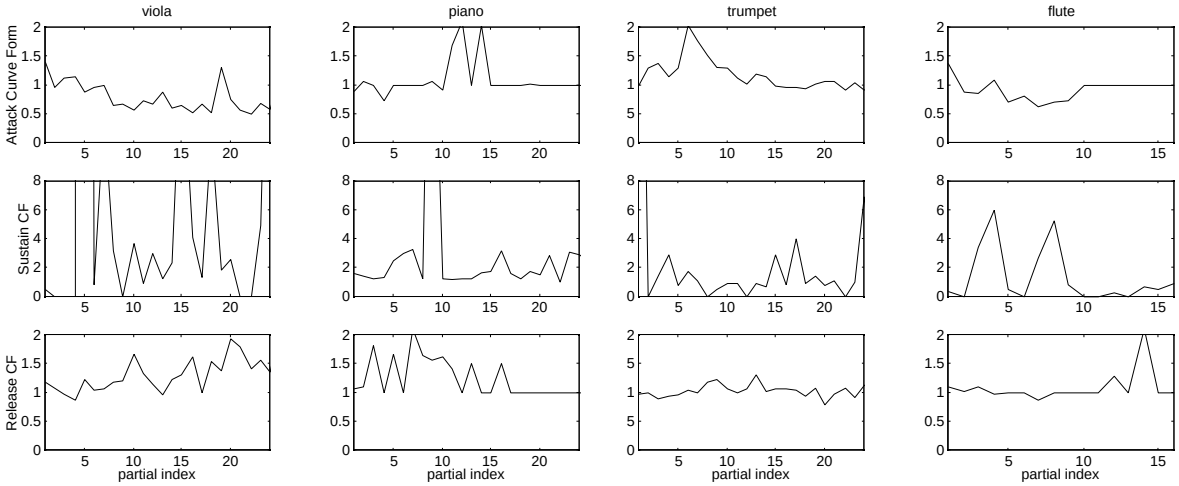


Figure 5.13. Curve form values for the slope analysis. Attack (top), sustain (middle) and release (bottom). Notice the different y scale for the sustain curve form.

The curve forms are close to linear, sometimes slightly logarithmic or exponential. It is interesting to observe the evolution of the curve form of the attack for the viola, the trumpet and the flute, which changes shape from quite logarithmic to quite exponential. The reason for this is not clear. It doesn't look very correlated with the amplitude percents in figure 5.9, and although it seems rather correlated with the envelope times in figure 5.8, it seems that the higher partials have a more exponential behavior than the lower partials.

5.4. Reconstruction of the Envelope

The envelope can now be recreated by concatenating the envelope segments with the analyzed envelope times, percents and curve forms.

The envelope consists of five elements, the start segment, attack segment, sustain segment, release segment and end segment. All envelopes start and end at zero amplitude by default.

The recreated envelopes of the fundamental of the viola, the trumpet, the piano and the flute are shown in figure 5.14 for the percent (top) and slope (bottom) based envelope analysis. The envelope split points are marked with plus signs in the plots.

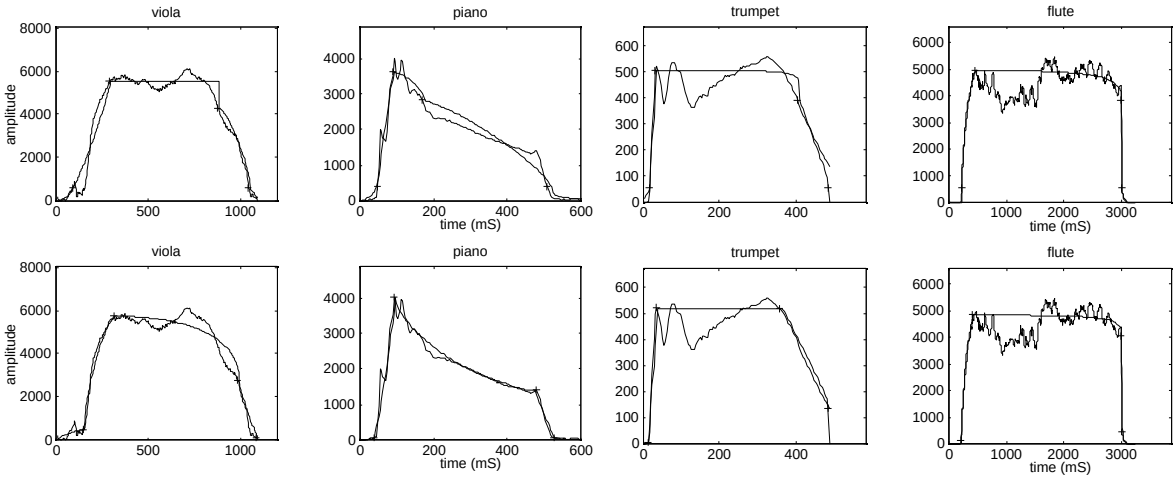


Figure 5.14. Original and percents (top) and slope (bottom) fundamental envelope for four sounds.

It can clearly be seen in the piano envelopes that the slope method gives the correct envelope break times. The percent method missed the attack of the viola due to noise preceding the attack, and especially the release of the piano, since the decay is not analyzed well with the percent method. The trumpet end of release percent is rather high, since the trumpet has been cut off before the end of release.

5.5. Recreation of the Additive Parameters

The additive parameters can be recreated from the envelope parameters, if each partial envelope is multiplied by the maximum amplitude of that partial, as found in the spectral envelope. The frequencies are set to the mean of the original frequencies.

The original, percent and slope additive parameters can be seen for the flute in figure 5.15, for the piano in figure 5.16, for the trumpet in figure 5.17 and for the flute in figure 5.18. The left plot is the original, the middle is the percent parameter plot and the right plot is the slope parameter plot. The frequencies of the reconstructed additive parameters are static with the value of the mean of the original analyzed frequencies.

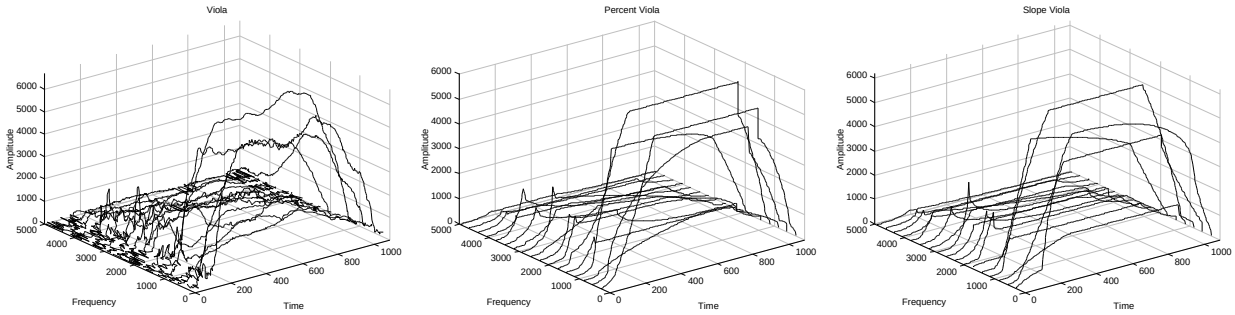


Figure 5.15. Viola additive parameters. Original (left), percent-based (middle) and slope-based (right).

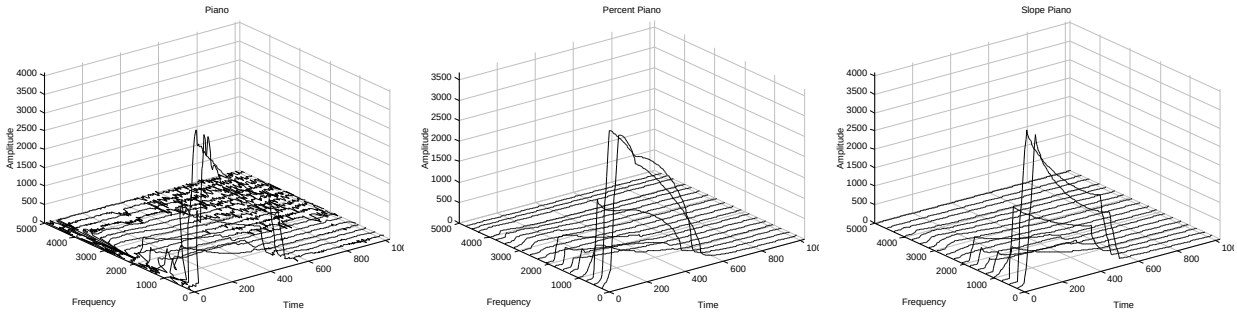


Figure 5.16. Piano additive parameters. Original (left), percent-based (middle) and slope-based (right).

Although it is difficult to distinguish all details here, it is quite obvious that the slope envelope time analysis improves the reconstruction of the additive parameters significantly. The slope-based additive parameters do not have the sharp edges that the percent-based additive parameters have where no segment change should happen. The slope-based additive parameters also fall closer to the split points in many occurrences, thereby allowing a better curve fit.

The piano additive parameters look much closer to the original envelope for the slope-based analysis than for the percent-based analysis.

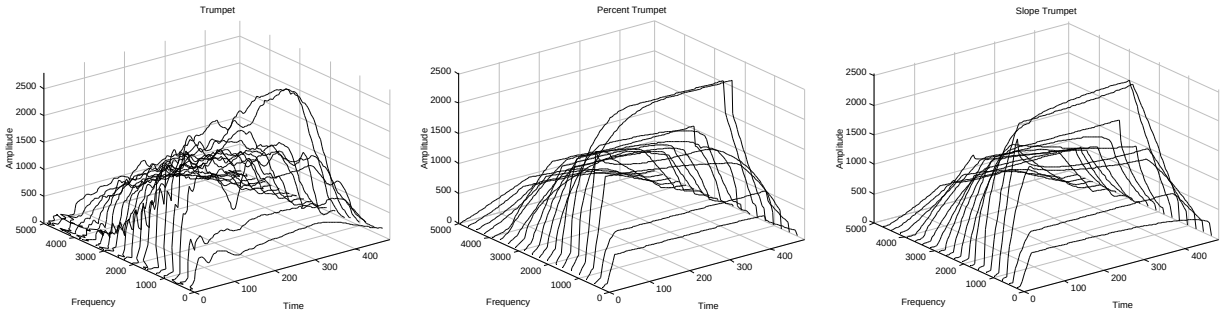


Figure 5.17. Trumpet additive parameters. Original (left), percent-based (middle) and slope-based (right).

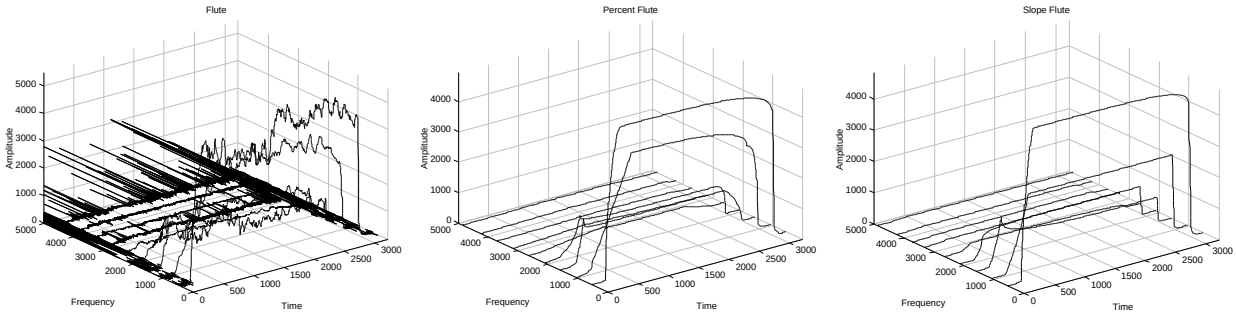


Figure 5.18. Flute additive parameters. Original (left), percent-based (middle) and slope-based (right).

Obviously, the best method for judging a model of a sound is to recreate the sound by the parameters of the model chosen. The resynthesis of the viola, trumpet, piano and flute sounds are very close to the original for both the percent and the slope envelope analysis. The main omission is the noise of the flute, and a general lack of presence in all sounds.

Even though the percent envelope time analysis performs worse than the slope analysis, this fact is partially masked by the curve form. Nevertheless, careful listening reveals the artificial character of the sounds of the percent method, especially in the attack of the flute. This artificial character of the sound was not discernible with the slope analysis sounds.

Although visually the percent method performs worse in the release, the informal listening tests mostly revealed problems in the attack.

5.6. Envelope Sharpening

Since the extraction of amplitude envelopes often is the result of an analysis using a window function, which introduces averaging, the envelope can be assumed to be less ‘sharp’ than the real envelope. A method for sharpening the envelopes would potentially restore the envelopes to the original form. In image processing, a deblurring technique has been used for some time [Hummel *et al.* 1987], [Kimia *et al.* 1993], [Haar Romeny *et al.*

1994], [Mair *et al.* 1996]. It is based on the assumption that images are blurred by a convolution with a gaussian. The deblurring is done by multiplying the Fourier domain signal by an inverse gauss. This increases the high frequency content of the signal. Although the method is unstable under bad conditions, it has been tested with some success in image processing.

Initial experiments with deblurring of the envelope are promising. It seems that this method can be used not only in the estimation of envelope times, but also in the modification of the additive parameters, perhaps permitting a better synthesis of fast transients. Furthermore, related methods have been used in time/frequency analysis [Gonçalvès *et al.* 1998].

5.7. Conclusion

A model for the amplitude of the additive parameters of a quasi-harmonic sound has been presented. It is based on four envelope time/value pairs and corresponding curve forms.

This model represents a ‘clean’ additive parameter set well. By clean is meant that the frequencies are static and noise-less, and that the amplitudes are noise-less. Furthermore it is assumed that there is no vibrato, glissando or tremolo.

This work presents a new envelope estimation method based on the analysis of the slope of the envelope. This method presents significantly better results than a simpler percent-based model. The estimation of important timbre attributes, such as the attack and release times, is improved, and the resynthesis of sounds from the slope analysis is better than the sounds from the percent analysis.

The introduction in this work of a simple intuitive envelope model with variable split-point amplitudes models both sustained and decaying sounds.

The envelope model with slope analysis of the envelopes can thus be said to model satisfactorily a harmonic noiseless sound with no vibrato or tremolo. This envelope model is used in the HLA model in Chapter 6, which also introduces a model of the noise and irregularities of the envelopes.

Chapter Six

6. High Level Attributes

In this chapter, the additive parameters found by the analysis presented in Chapter 4 are further modeled. The assumption that a musical sound fits the envelope model introduced in Chapter 5, with silence, attack, sustain or decay, release and again silence, is used.

The High Level Attribute (HLA) model is created by extracting meaningful parameters from the very large additive parameter data set. The parameters of the HLA model can be divided into amplitude envelope, spectral envelope, frequency, and noise. It can be used to resynthesize sounds, morph between sounds, or understand timbre features of a sound. The sounds created from the HLA model are of good quality. The HLA model is the main timbre model in this work, although it still has too many parameters to permit a visualization of the timbre attributes of several sounds. This problem is addressed in the following chapters.

6.1. Introduction

The additive parameter description is a good model of quasi-harmonic sounds, but it has a very large, non-intuitive parameter set.

The High Level Attribute term was coined in [Serra *et al.* 1997]. They state, “from ... sinusoidal plus residual model higher level attributes such as: pitch, spectral shape, vibrato, or attack characteristics can be extracted”.

The HLA model introduced in this work can be seen as a data reduction of the additive parameters. Other data reductions techniques include the Group Additive Synthesis [Kleczowski 1989], [Eaglestone *et al.* 1990], [Cheung *et al.* 1996], where similar partials are grouped together to improve efficiency. Other means of improving efficiency in the resynthesis of the additive parameters include the multirate additive synthesis [Phillips *et al.* 1996] or the inverse FFT synthesis [Rodet *et al.* 1992].

The HLA model resembles a new class of speech coders, called sinusoidal coders [Gersho 1994], and especially the hybrid harmonic coding algorithms [Marques *et al.* 1994], although the HLA model is designed especially for isolated musical sounds.

Other uses of related methods include [Tellman *et al.* 1995] who uses envelope time points to morph between different musical sounds. Back in 1966 [Strong *et al.* 1966] synthesized wind instruments with a combination of spectral and temporal envelopes. [Rodet *et al.* 1987] use spectral envelopes as a filter with different source models, including the additive model.

The HLA model models each partial in a few pertinent parameters: the amplitude envelope, the spectral envelope, frequencies, and noise parameters. The maximum amplitude defines the spectral envelope, the mean frequency defines the frequency of each partial, the envelope is based on an attack-sustain-release, or attack-decay-release model presented in Chapter 5, and finally the irregularity of the partial amplitude and frequency models the noise of the sound. The HLA model has a fixed parameter size, dependent only on the number of partials, and the parameters of the HLA model have an intuitive perceptive quality.

The HLA model can be used to resynthesize the sound, with some or all of the parameters of the model. In this way, the validity of each parameter of the HLA model can be verified.

The chapter starts with an overview of the additive analysis in section 6.2, then the spectral amplitude model is presented in section 6.3, the frequency model is presented in section 6.4. An overview of the envelope model is presented in section 6.5 and the noise model is presented in section 6.6.

Finally a proposed visualization of the HLA parameters is introduced in section 6.7, the recreation of the additive parameters is presented in section 6.8 and the chapter ends with a conclusion.

6.2. Additive Parameter Analysis

The additive parameters are analyzed by the LTF analysis method presented in the analysis chapter.

The additive parameters are smoothed over one period of each sound and only pseudo-harmonic partials are saved. The good timing resolution of the LTF analysis permits a better analysis of fast transients, such as the attack of the piano, but it also models better the noise of the sound. This is used in the noise model, which models the noise as the irregularity on the amplitude and frequency of the partials.

6.3. Spectral Envelope

The spectral envelope is defined in this work as the maximum amplitude of each partial.

The spectral envelope is very important for the perceived effect of the sound; indeed, the spectral envelope alone is often enough to distinguish or recognize a sound.

This is especially true for the recognition of vowels, which are entirely defined by the spectral envelope.

Nevertheless, the spectral envelope alone is not enough to recreate any sound with realism.

The spectral envelopes for four musical instrument sounds are plotted in figure 6.1. The y-axis is the amplitude, where the amplitude scales are the same for all four sounds, and the x-axis is the partial index, which is proportional to the frequency.

Each spectral envelope has a distinct look, although there is almost no formant structure.

The viola spectrum has a very irregular shape, which is probably caused by the different damping of the modes [Benade 1990]. The piano has some missing (weak) partials, which are missing because these modes are annulled due to the hammer impact position [Hall *et al.* 1987]. The trumpet has the typical rising spectrum for the low partials [Benade 1973]. The flute is a higher pitched sound and thus naturally it has less energy in the higher partials.

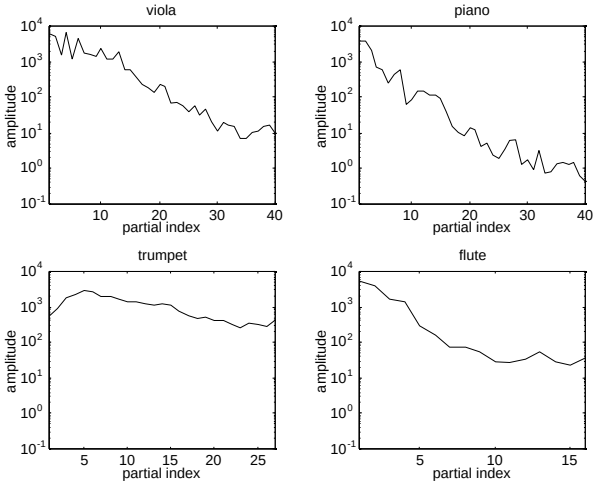


Figure 6.1. Spectral Envelope for the viola, the piano, the trumpet and the flute.

The flattening of the spectral envelope for the high partial index is probably due to background noise.

6.4. Frequency

The frequency of each partial is modeled as the mean of the frequency for the sustain part. Most sustained instruments are supposed to be perfectly harmonic. The piano, in contrast, has inharmonic partial frequencies due to the stiffness of the strings [Fletcher 1964].

The frequencies are best viewed divided by the partial index as seen in figure 6.2.

The frequencies divided by the partial index have a constant value for perfectly harmonic sound, if the partials contain only the harmonic overtones, as is the case here. The degree of inharmonicity for the piano is easy to see. Notice the y-axis scale for the piano. The high order partial frequencies can be misjudged due to the presence of noise, and should not be relied upon.

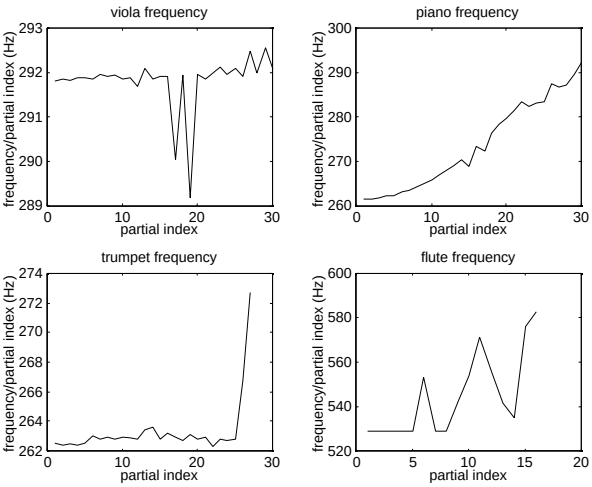


Figure 6.2. Frequency divided by the partial index for the viola, the piano, the trumpet and the flute.

The presence of inharmonicity in the piano certainly adds a flavor to the sound, and it is necessary to keep the frequency of each partial, instead of assuming clean harmonic frequencies.

6.5. Envelope

The envelope of each partial is modeled in five segments, a start and end segment, supposedly close to silent, and an attack, sustain segment and release segment. Thus, there are 6 amplitude/time split points, where the first is (0,0) and the last amplitude also is zero, since all partials are supposed to start and end in silence. The amplitudes are saved as a percentage of the maximum of the amplitude, and the times are saved in mS. Furthermore, the curve form for each segment is modeled by a curve, which has an appropriate exponential/logarithmic form.

A further development of the envelope analysis can be found in Chapter 5. A short description of the method used to find the envelope times, percents and curve forms will nevertheless follow here. The slope method, which was developed in this work, is necessary for the proper estimation of the attack and release times.

6.5.1. Timing Analysis

The attack and release segments are found by searching the maximum and the minimum of the derivative of the amplitude of each partial, and the start and end of each segment is found by following the derivative until its absolute value is below the maximum times a threshold.

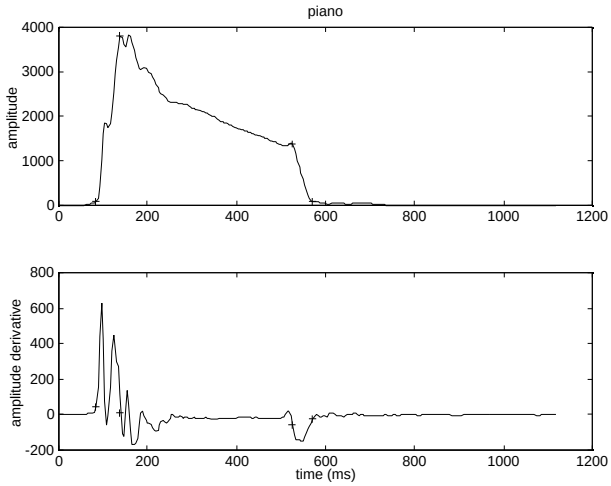


Figure 6.3. Envelope (top) and first derivative (bottom) times for the fundamental of the piano.

The method can be seen in figure 6.3. Here the amplitude of the fundamental of the piano is plotted with its first derivative. The '+' depict the split points. As can be seen, the start of release threshold is larger than the other three thresholds.

This is so the release of the piano and other plucked/damped instruments, which have a decay/release envelopes can be properly analyzed.

The envelopes are in general too noisy for this analysis, so it is done on a heavily smoothed envelope.

The smoothing ensures that the local minimum of the derivative, as seen in figure 6.3, do not ruin the analysis.

The envelope times are then followed from the smoothed to the unsmoothed envelope by a method inspired by the scale-space theory [Lindeberg 1996].

This method succeeds, as can be seen in figure 6.3, to find the proper release time for the piano sound. The attack has also been properly estimated, even though the derivative of the amplitude decreases to below zero in the middle of the attack. These fast variations are not present in the smoothed envelope, which is used for the first estimation of the envelope times. Again, see the envelope modeling in Chapter 5 for more details.

The perceptually most important envelope parameters seem to be the attack and release times. These are easily calculated from the difference between the absolute times, and they are shown in figure 6.4. The top plots are the attack times, and the bottom plots are the release times.

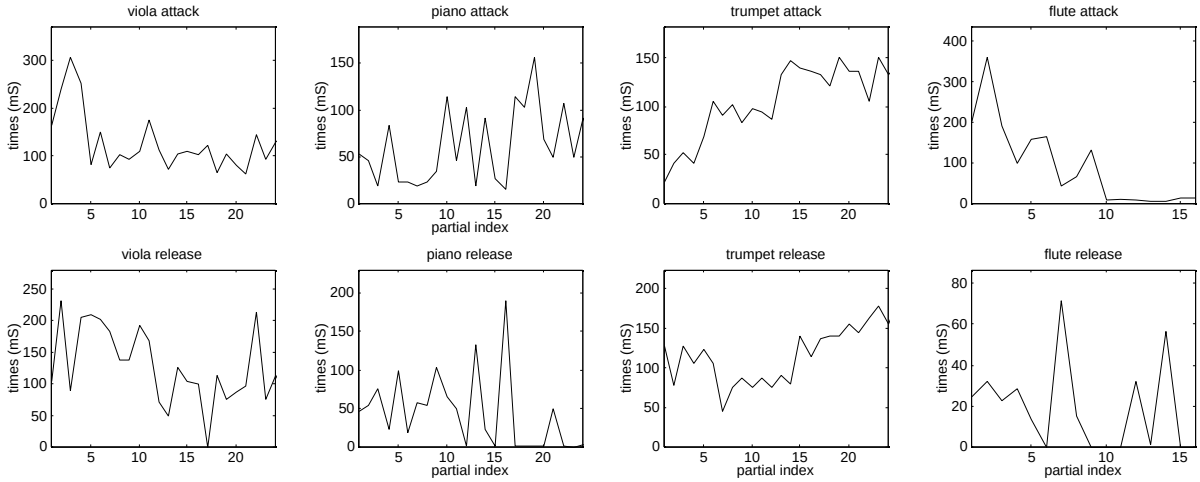


Figure 6.4. Attack (top) and release (bottom) times for the viola, the piano, the trumpet and the flute.

The estimations of the attack and release times are not noiseless, but some observations can still be made. The release times are normally in the same range as the attack times, except for the flute, which has very fast release times. The viola attack and release seems to decrease with frequency, from about 200 mS to 100 mS. The piano has a fairly constant attack and release time of about 50 mS. The trumpet attack and release times increase with frequency, from around 50 mS to 150 mS. The flute attack seems to decrease from almost 300 mS to zero for the high partials. These can be misjudged because of the additive noise present in the flute sound.

The attack and release times seem rather reliable, and the behavior of these times gives a lot of information about the musical sound.

6.5.2. Curve Form Analysis

The curve form of each segment is modeled by a curve with appropriate logarithmic or exponential form. The curve form is usually close to linear, sometimes logarithmic and sometimes exponential.

The curve used for the modeling of the envelope for one segment is

$$Curve_s = v_0 - (v_1 - v_0)(1 - (1 - x)^n)^{\frac{1}{n}} \qquad \textbf{(6.1)}$$

and the curve form value *n* is found by minimizing the squared error,

$$Error = \sum_{t=1}^{N} \left(Curve_t - Envelope_t\right)^2 \qquad \textbf{(6.2)}$$

The resulting envelopes can be seen in figure 6.5 (dotted) with the original envelopes for the fundamental of four instruments. It is interesting to see the general form of the viola, the trumpet and the flute, where the envelope initially rises to a value close to the maximum, but then weakens slightly and then rises to the maximum value. This seems to support the quietest point of the envelope introduced in [Tellman *et al.* 1995], although this parameter has not been judged perceptually important enough to be included here.

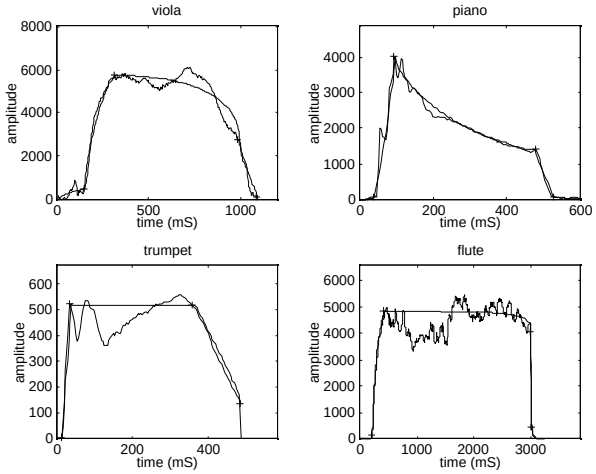


Figure 6.5. Original and recreated fundamental envelope for the viola, the piano, the trumpet and the flute.

The piano has an irregularity in the attack. This irregularity is also present the attack of most of the trumpet partials. This effect has not been found in the literature, and it has not been found necessary to model it in the envelope model, but it is instead modeled in the noise in the next section.

6.6. Noise

Although the recreated envelopes in figure 6.5 have the general shape of the original envelope, it is easy to see that there is a great deal of irregularity left, which is not modeled. The same holds true for the frequency. The noise on the amplitude envelope is called shimmer, and the noise on the frequency is called jitter [Richard *et al.* 1996]. Shimmer and jitter are modeled for the attack, sustain and release segments. The noise is supposed to have a Gaussian distribution; the amplitude of the noise is then characterized by the standard deviation. The frequency magnitude of the noise is modeled, as is the correlation between the shimmer and jitter of each partial and the fundamental.

Other noise models of musical sounds include the residual noise in the FFT [Serra *et al.* 1990], [Møller 1996] and the random point process model of music noises [Richard *et al.* 1993] or speech noise [Richard 1994], [Richard *et al.* 1996]. Models of noise on sinusoidals include the narrow band basis functions (NBBF) in speech models [Marques *et al.* 1994]. In music analysis, [Fitz *et al.* 1995] have introduced the bandwidth enhanced sinusoidal modeling. Both models model only jitter, not shimmer. Other analysis of the noise, and irregularity of the music sounds include the analysis of aperiodicity [McIntyre *et al.* 1981], [Schumacher *et al.* 1990], and the analysis of higher order statistics [Dubnov *et al.* 1996], [Dubnov *et al.* 1997].

6.6.1. Distribution of Partial Noise

Shimmer and jitter are supposed to be normally distributed, and the amplitude is calculated by the standard deviation. Shimmer is correlated with the maximum amplitude of the partial, whereas jitter is correlated with the mean of the frequency of the partial. The shimmer and jitter standard deviations are therefore modeled as a percentage of the value of the amplitude curve model and the mean frequency,

$$\sigma_{shimmer} = std\left(\frac{a_t - c_t}{c_t}\right) \quad \textbf{(6.3)}$$

$$\sigma_{jitter} = std\left(\frac{f_t - \dot{f}}{\dot{f}}\right) \quad \textbf{(6.4)}$$

a and *f* are the time-varying amplitudes and frequencies of the partial, $|\dot{f}|$ is the mean frequency and c_t is the curve found by the envelope model. If the noise magnitude has a peak above zero frequency, it is assumed to be vibrato, or tremolo, and removed before the std calculation.

Resynthesis of the sound with either shimmer or jitter makes it possible to evaluate the importance of each parameter, and even though shimmer and jitter each add a quality to the sound, shimmer seems more important, at least for the flute sound. This is probably because the amplitude model is too simple for the actual amplitude, especially for long sounds, and not necessarily because shimmer is more perceptible than jitter.

6.6.2. Spectrum of Partial Noise

The spectrum of shimmer and jitter is supposed to be band-limited, and is modeled as white noise passed through a single-tap recursive filter,

$$noise_t = noise_t - a \;\; noise_{t-1} \qquad \textbf{(6.5)}$$

The magnitude response of this filter is [Steiglitz 1996],

$$|H(\omega)| = \frac{1}{\sqrt{1 + a^2 + 2a \;\cos(\omega)}} \qquad \textbf{(6.6)}$$

The filter coefficient *a* is found by a least-squares fit to the original noise frequency magnitude response.

The influence of the standard deviation and filter coefficients of the shimmer and the jitter can be seen in figure 6.6 to figure 6.9. The power spectral density (PSD) [Press *et al.* 1997] estimation of a single sinusoidal with variable shimmer and jitter parameters is plotted.

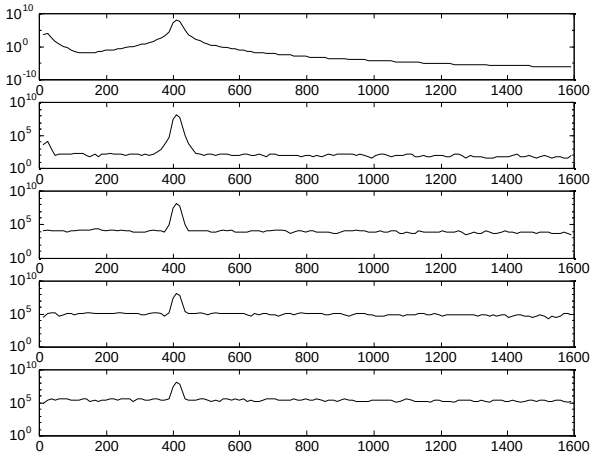


Figure 6.6. Influence of the std of the shimmer from 0 (top), 0.01, 0.1, 0.3 and 0.5 (bottom). No jitter, filter coefficient of the shimmer is -0.5.

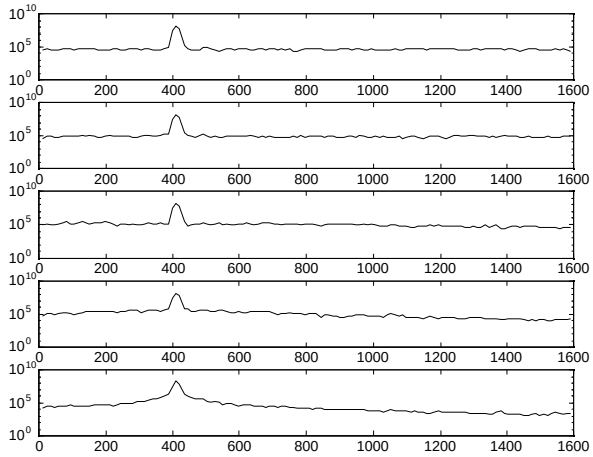


Figure 6.7. Influence of the filter coefficient of the shimmer, with std 0.3, from 0 (top), -0.3, -0.7, -0.9, -0.99 (bottom), no jitter.

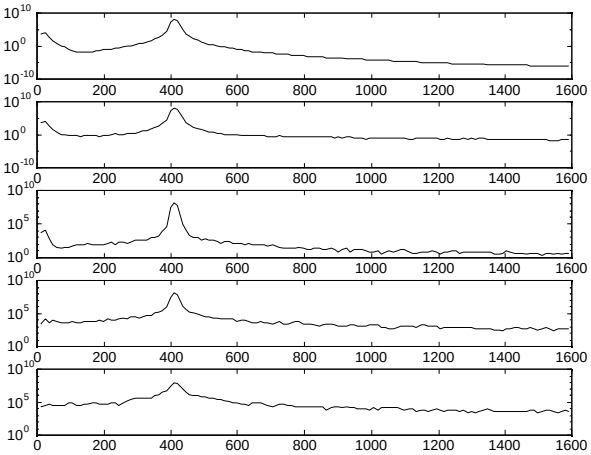


Figure 6.8. Influence of the std of the jitter from 0 (top), 0.01, 0.1, 0.3 and 0.5 (bottom). No shimmer, filter coefficient of the jitter is -0.5.

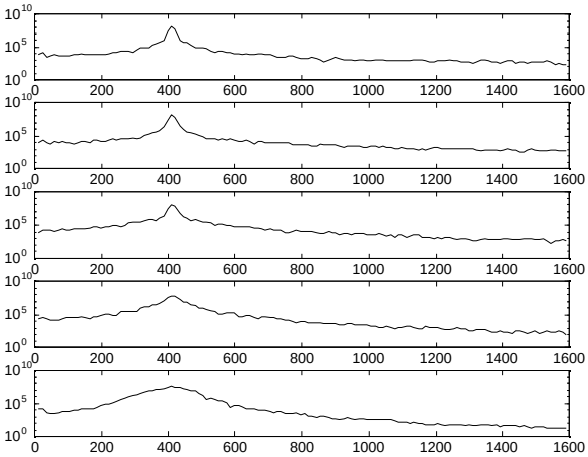


Figure 6.9. Influence of the filter coefficient of the jitter from 0 (top), -0.3, -0.7, -0.9, -0.99 (bottom). No shimmer, std of the jitter is 0.1.

Shimmer is an additive component in the frequency domain, whereas jitter increases the bandwidth of the sinusoidal.

When the filter coefficient decreases towards –1 the bandwidth of the sinusoidal decreases. A filter coefficient of zero gives band-pass noise.

Shimmer has an additive noise quality; the std increase gives the effect of more noise, and the filter coefficient decrease (toward -1) gives the effect of more band-pass noise.

For jitter, the std increases the noise, which has a different quality than shimmer, more band-passed it seems, and the noise quality of the filter coefficient decrease goes from an additive noise to modulating frequency, ending in low-frequency jitter modulation.

6.6.3. Correlation of Partial Noise

The correlation of the shimmer and the jitter is calculated between each partial and the fundamental. This is done to separate correlated noise from non-correlated noises. Other, more elaborate models, such as the phase coupling between partials, have not been tested.

The standard deviation, filter coefficient and the correlation for the jitter of the four test sounds can be seen in figure 6.10. The standard deviation of the jitter, normalized with the frequency, is shown on top, the filter coefficients in the middle, and the correlation in the bottom plot. All y scales have been normalized to facilitate comparison.

The shimmer parameters are shown in figure 6.11 with the same disposition of the parameters.

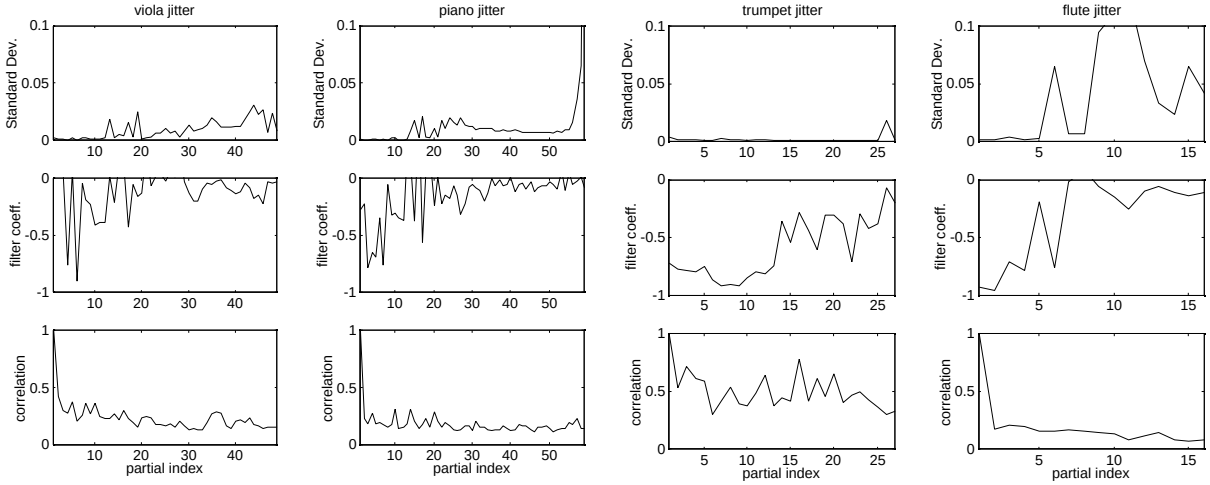


Figure 6.10. Partial frequency noise (jitter) parameters. Standard deviation (top), filter coefficients (middle) and correlation (bottom) for the viola (left), the piano, the trumpet and the flute (right).

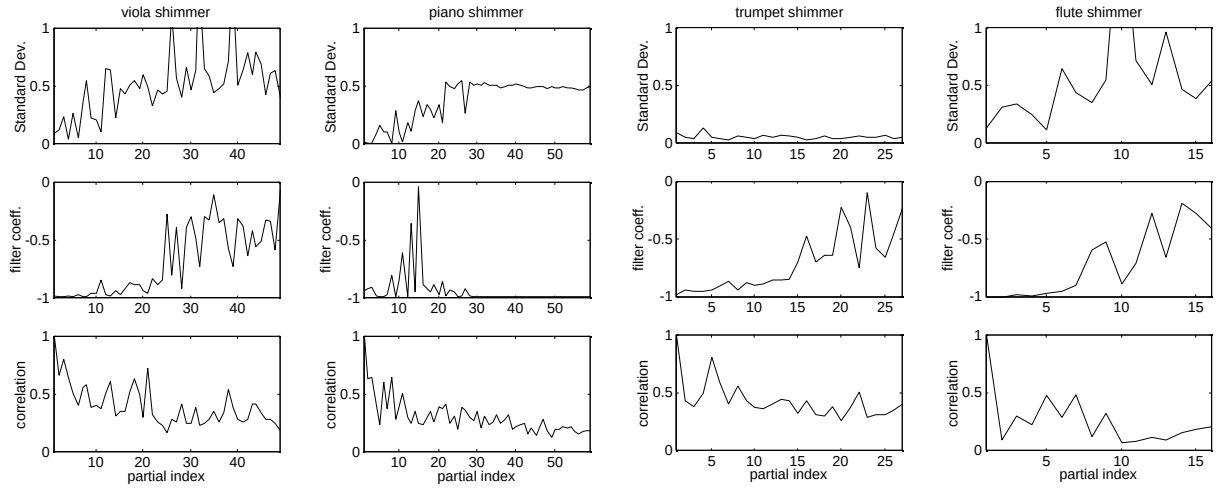


Figure 6.11. Partial amplitude noise (shimmer) parameters for the viola (left), the piano, the trumpet and the flute (right).

The jitter parameters have some common features for all sounds. The jitter std is very low for the low strong partials, and rising for the weak upper partials. The filter coefficients are approaching zero for the high partials. The correlation is falling slowly with partial index.

There seems to be a noticeable similarity between the two string sounds, viola and piano, and the two wind sounds, trumpet and flute. Shimmer generally has a lower filter coefficient, and thus more low-frequency energy, which seems to be caused by mismatch of the simple envelope approximation. The shimmer std is much higher than the jitter std, and rising with partial index, except for the trumpet, which has a fairly stable shimmer std. The shimmer correlation does not seem significantly higher than the jitter correlation, as should be expected.

6.6.4. Resynthesis of Noise

The shimmer and jitter noise of each partial is the sum of two filtered noises, one independent ns^i, and one common to all partials, ns^c. In order to avoid abrupt changes in the noise, it is recreated using three envelopes.

The attack noise envelope is a ramp going linearly from zero at the beginning of the attack, to one at the middle of the attack and again to zero at the end of the attack.

The release noise envelope is similar, whereas the sustain noise envelope is one in all of the sustain region going linearly to zero in the middle of the attack and the release.

The noise envelopes can be seen in figure 6.12. The attack and release noise envelopes are zero at the split points. This makes sense, since the error also is zero at the split points.

The sustain noise, which generally is much lower than the attack and release noises, is prolonged into the attack and release segments to avoid abrupt changes.

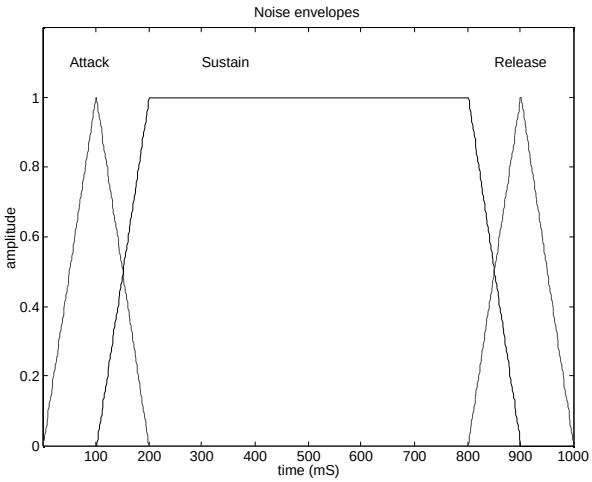


Figure 6.12. Noise envelopes. Attack and Release (dashed) and Sustain (solid).

The total shimmer or jitter for partial *k* and segment *s* is,

$$ns_{s,k} = envelope_{s,k}(t)\ \ \sigma_{s,k}\ \ filter_{s,k}((1-c_k)\ ns^i_{s,k} + c_k\ \ ns^c_s) \qquad \textbf{(6.7)}$$

where c_k is the correlation coefficient for the partial *k* and $\sigma_{s,k}$ is the standard deviation for segment *s* and partial *k*. The three shimmer segments are now added to the clean envelope at the appropriate times, and the three jitter segments are added to the static frequency for the partial. The start and end segments do not have any noise.

6.6.5. Noise Conclusion

The jitter and shimmer are here modeled by a normal distribution with mean zero. Furthermore the spectrum of the noise is modeled using a simple recursive filter. This seems to be sufficient in many situations, but it leaves room for improvements. Although the noise effectively seems to be gaussian, it might not always have the same skewness, or kurtosis [Press *et al.* 1997]. Skewness is a measure of the asymmetry of the distribution,

and kurtosis is a measure of the peakedness of the distribution. Experiments with higher order moments have not been performed here. [Ystad 1998] finds the distribution of the source of the flute to be exponential, and offers an algorithm for the construction of an exponentially distributed noise. [Dubnov *et al.* 1997] finds the phase coupling to be an important characteristic of musical instruments. This might be an important addition to the noise model. It is somehow difficult to judge the quality of the noise of the resynthesis here, since the amplitude model is not assumed to give an identical shape, and much of the shimmer noise, or irregularity, originates from this fact.

6.7. HLA Visualization

The HLA set can be divided into 4 groups, the spectral envelope, the frequencies, the envelope and the noise parameters. The envelope group can be further divided into the envelope timing, the envelope percents, and the envelope curve forms. The noise can be divided into the shimmer and jitter standard deviations, the shimmer and jitter filter coefficients, and the shimmer and jitter correlation.

In total, there are 10 groups, which can be plotted in one figure in 5 rows and 2 columns. The left column has from the top to the bottom the spectral envelope, the frequencies divided by the partial index, and envelope timing, the envelope percents and the envelope curve forms. The right column has from the top to the bottom the shimmer standard deviation, the jitter standard deviation, the shimmer filter coefficients, the jitter filter coefficients and the shimmer and jitter correlation.

In figure 6.13, figure 6.14, figure 6.15 and figure 6.16 are shown the complete HLA set for the 4 sounds, viola, piano, trumpet and flute. To improve visibility, only the 16 first partials are plotted.

The spectral envelope (top left) and the frequencies (second from top left) have only one curve each. The spectral envelope is plotted in the log domain.

The envelope timing (third from top left) has four curves, the start of attack time ‘o’, the end of attack time ‘*’, the start of release time ‘x’, and the end of release ‘+’. The percents (fourth from top left) also have four curves with the same symbols, the curve forms (bottom left) have 5 curves, the start curve ‘+’ the attack curve ‘o’, the sustain curve ‘*’, the release curve ‘x’ and the end curve ‘.’. For the sake of clarity, the start curve and the end curve are dotted.

The noise attributes generally have 3 curves, attack ‘o’, sustain ‘*’, and release ‘x’, with the exception of the noise correlation, where the shimmer is ‘o’ and the jitter is ‘*’. The shimmer std is plotted top right, the shimmer filter coefficient is second from top right, the jitter std is third from top right, the jitter filter coefficients are fourth from top right, and the noise correlations are plotted bottom right.

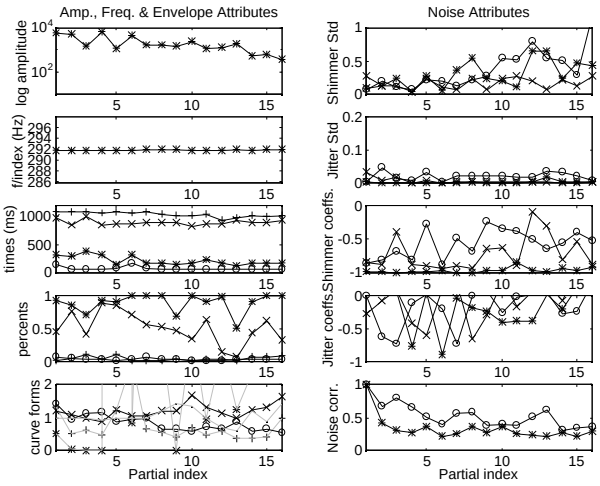


Figure 6.13. Complete HLA set for the viola sound.

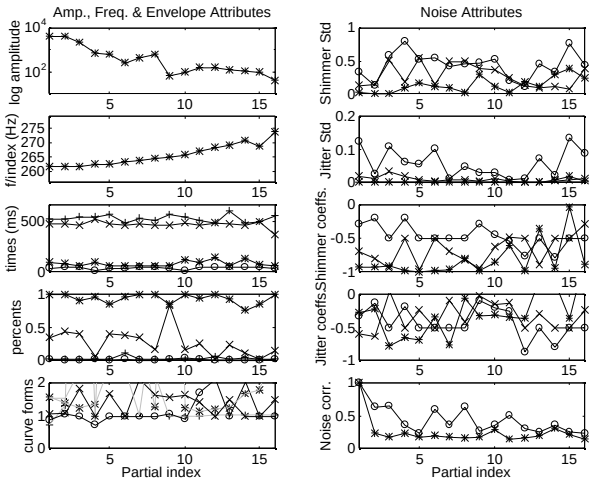


Figure 6.14. Complete HLA set for the piano sound.

The viola has a quite irregular spectral envelope, with much energy in the high partials. The frequencies seem perfectly harmonic. The estimation of the envelope parameters seems stable; the attack and release times decrease with frequency, as does the end of release percents. This is an indication of the faster decay of the higher partials. The attack curve form seems to change from logarithmic for the low partials, to exponential for the high partials.

The shimmer std increases with frequency, whereas the jitter std is rather stable. The shimmer filter coefficient is close to -1 for the fundamental, rising towards zero for the high partials. The filter coefficients are higher for the high partials, it seems, because of the shorter duration of these partials. The short partials have by definition a better curve fit, which translates into a low shimmer filter coefficient. The jitter filter coefficients are more stable. The correlation is decreasing slightly for both the shimmer and jitter for the viola sound.

The piano spectral envelope has a weak formant at the eighth partial. The frequencies are stretched, the envelope times rather stable. The end of release percents is much lower than the end of attack percents due to the decay slope of the envelope and falling with the partial index. The attack and release shimmer is rather high, with a relatively low filter

coefficient for the attack segment, indicating fast irregularities. The jitter correlation is lower than the shimmer correlation for the piano.

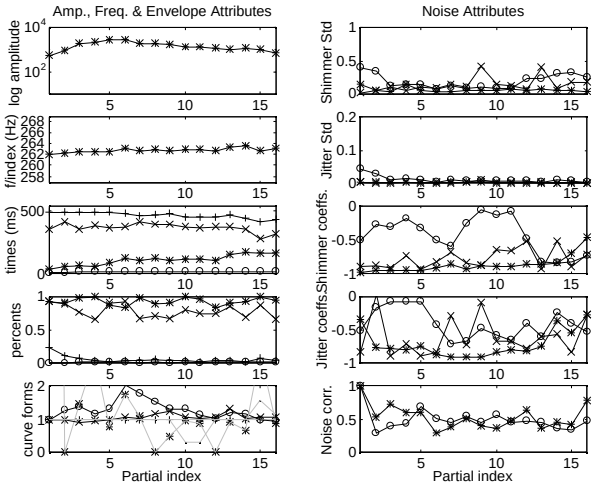


Figure 6.15. Complete HLA set for the trumpet sound.

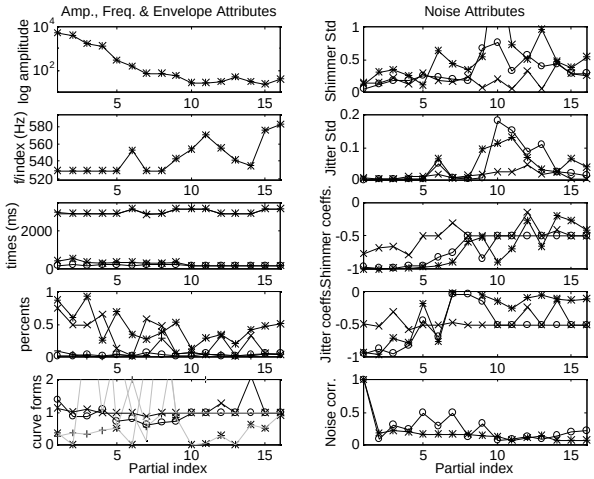


Figure 6.16. Complete HLA set for the flute sound.

The trumpet has a strong formant at around the fifth partial. The frequencies are harmonic. The attack and release times increase considerably with the partial index for the trumpet. This is part of what gives the characteristic trumpet sound [Risset 1965]. The percents are very close to one for all partials, which proves the validity of the envelope times. There is not much noise in the trumpet partials, although the attacks of the first few partials have more high frequency noise, which can be seen by the high std of the first partials and the relatively high values of the attack filter coefficients.

The flute has few strong partials. The frequency of the strong partials is harmonic, and the other frequencies are noisy. The envelope times are difficult to see, since the flute sound is so long. The percents decrease with frequency. There is rather much shimmer in the flute sound, indicating more additive noise, although the shimmer filter coefficient is close to -1, which is more an indication of envelope curve misfit. The jitter correlation is rather low for the flute.

In conclusion, the HLA parameters give important information about the sound they derive from. The spectral envelope, the length, the attack and release characteristics and the noises are easily seen in the HLA visualization, or compared with other sounds.

6.8. Recreation of the Additive Parameters.

The additive parameters are recreated from the HLA by first creating clean amplitudes with the envelope model presented in section 6.5 and frequencies with the frequency

model presented in section 6.4, and then adding the noise on the parameters as explained in paragraph 6.6.4. The noise is added in the attack, sustain and release segments. There is no noise in the start or end segments, since they are assumed to be silent.

The original and recreated additive parameters for 4 sounds can be seen in figure 6.17 (viola), figure 6.18 (piano), figure 6.19 (trumpet) and figure 6.20 (flute). The left plots are the original LTF analyzed additive parameters and the right plots are the HLA model recreated additive parameters.

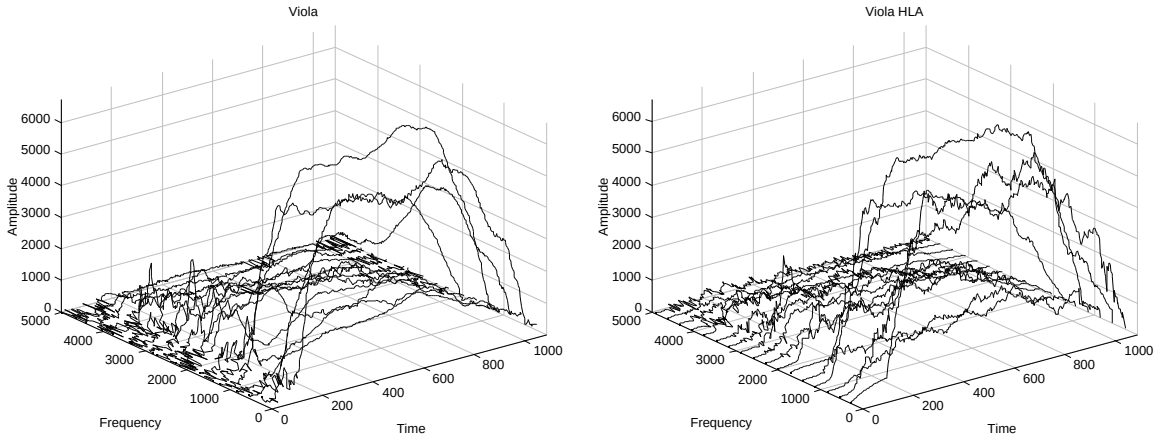


Figure 6.17. Original and MDA recreated additive parameters for the viola.

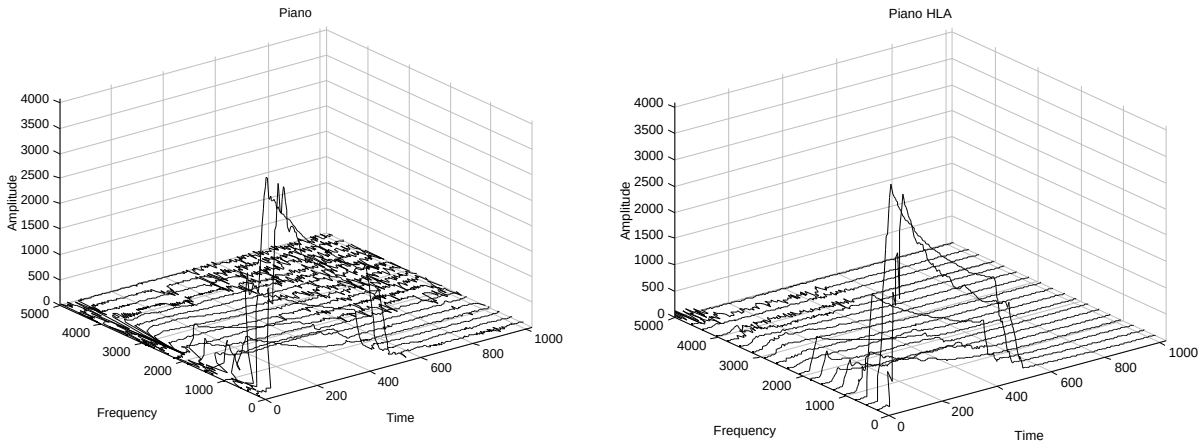


Figure 6.18. Original and MDA recreated additive parameters for the piano.

The visual shape of the additive parameters is well preserved, obviously without having an identical form. The noise part of the parameters is random, so it is never two times the same sound, or the same visual shape. Some differences in the noises are nonetheless clear. This might be explained by the simple filter model of the noise, or by an incomplete description of the noise distribution.

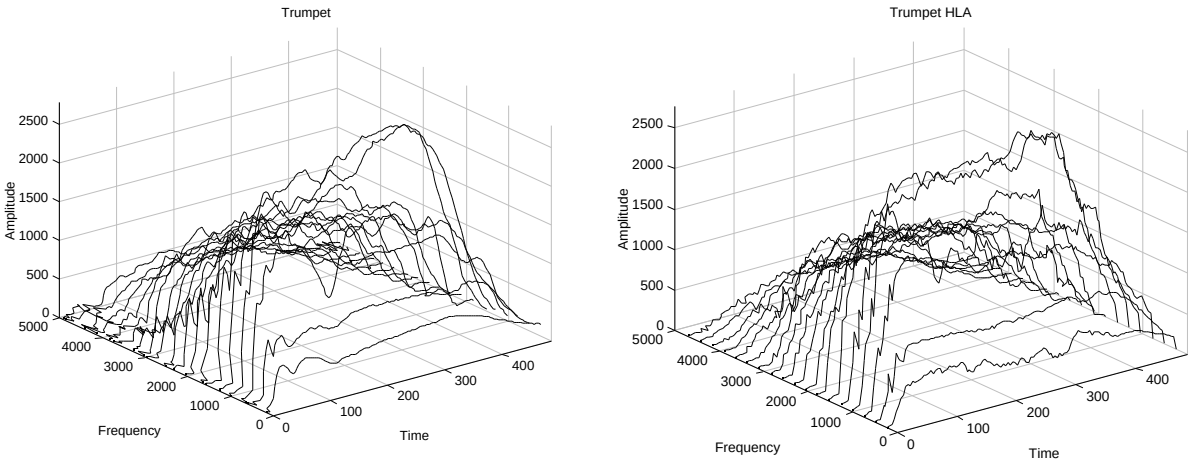


Figure 6.19. Original and MDA recreated additive parameters for the trumpet.

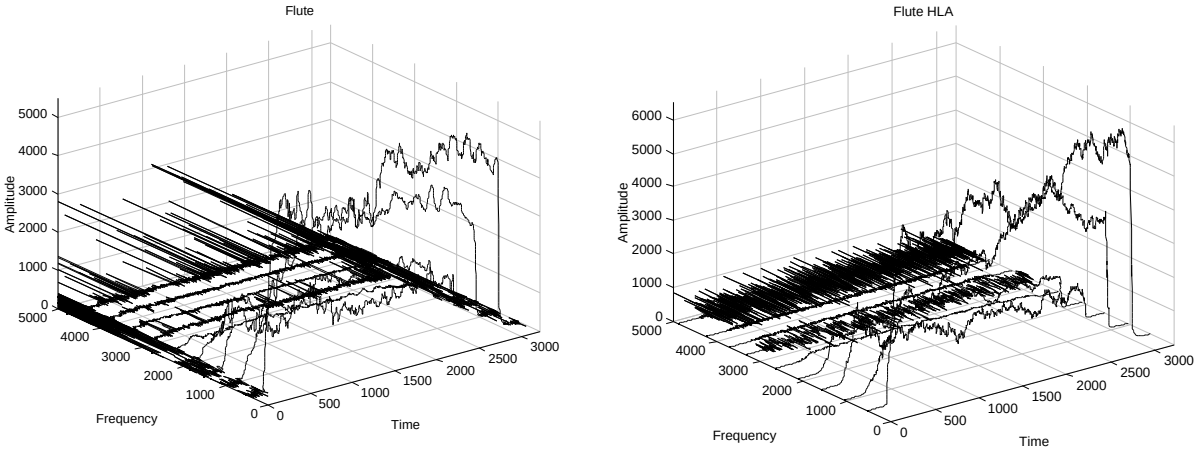


Figure 6.20. Original and MDA recreated additive parameters for the flute.

The resynthesis of the sounds from the HLA permits the evaluation of the validity of the HLA model. Generally, the HLA model is good for the synthesis of musical sounds. The pitch, loudness and duration are recreated flawlessly, as is most of the timbre attributes. The sounds are always identifiable although always different from the original sounds. The quality of the HLA model resynthesis is generally very good, and typical features, such as formants, noise, or sharp attacks are always present in the resynthesis.

Although the HLA model is sensitive to bad fundamental estimation and bad curve fitting, which tends to increase the noise factor [Marques *et al.* 1994], this does not seem to be a serious problem. The quality of the HLA is also degraded, if there is vibrato or tremolo in the original sounds. This translates into noise in the HLA model.

6.9. Conclusion

This work presents a new model of musical instrument sounds. The HLA model models the additive parameters in a few intuitive parameters: the spectral envelope, amplitude envelope, mean frequency, and noise.

The HLA model is well suited for isolated sounds. It does not model vibrato, tremolo or glissando, which are supposed to be user-provided, that is, performance expressions. Listening tests presented in Chapter 12 show that musical sounds are resynthesized well with this model. Furthermore, the HLA model parameters help in the understanding of timbre and the perceived difference of sounds. Important timbre cues, such as the spectral envelope, the envelope timing, and the noise are easily extracted and visualized from this model.

The HLA model, as implemented here, permits an automatic analysis/synthesis of musical sounds with a small parameter size. The fixed parameter size of the HLA model is helpful when comparing sounds, or when timbre morphing is performed. However, the HLA model does not permit the visualization of a single timbre attribute, such as the attack time, for many sounds. Therefore, the HLA model is further simplified in the next chapters.

Chapter Seven

7. Spectral Envelope Model

The spectral envelope is here defined to be the maximum amplitude of the quasi-harmonic partials of a sound. This chapter presents a model of the spectral envelope, based on some perceptually meaningful attributes. These attributes are calculated on the quasi-harmonic components of the original spectrum. In the reconstruction, a new spectrum is created with the same attribute values as the original spectrum.

This model, using perceptive attributes, is valid for non-formantic sounds. Initial listening tests have confirmed the validity of the model.

The purpose of this work is to create a stable analysis/synthesis method of the spectral envelope, using a few intuitive parameters.

When interpolating from one spectral envelope to another, the spectral envelope model permits in theory the displacement of important timbre features, instead of the lowering of one feature and the increasing of the other.

7.1. Introduction

This chapter models the spectral envelope of musical sounds. The spectral envelope is defined as the maximum of the amplitudes of the quasi-harmonic overtones. Based on the spectral envelope, a few perceptually important parameters can be calculated, and used for the subsequent recreation of a synthetic spectral envelope with the same perceptual timbre. Furthermore, these parameters can be calculated for a time-varying spectrum, and a synthetic spectrum can be recreated with the same time-varying perceptual values. The parameters of the spectral envelope model are the brightness, tristimulus, odd value, irregularity and maximum amplitude.

The spectral shape of a sound is often modeled by a source/filter strategy. This is the case when modeling speech [Klatt 1980], where the source generally is divided into a voiced part, which is defined by the dB/octave slope, and a noise part. The filter is generally a number of resonators, which corresponds to the formants of speech. More accurate source models for the voiced part of the speech have been introduced in for instance [Fant *et al.* 1985] and [Veldhuis 1998]. Often the filters are modeled by the linear predictive coding (LPC) [Rabiner *et al.* 1978]. Several papers use the spectral envelope as the filter part [Strong *et al.* 1966], [Rodet *et al.* 1987], [Rodet *et al.* 1992], [Horner *et al.* 1995].

In music research, the spectral envelope is often created by a non-linear function, such as frequency modulation (FM) [Chowning 1973], and a great number of similar techniques [Arfib 1978], [le Brun 1979], [Mitsuhashi 1982], [de Poli 1984], which, while generating complex spectra with low processor cost, generally lacked both analysis techniques, and intuitive control. Many attempts have been made to match the parameters of a processor-effective algorithm, such as the FM, to the parameters of an acoustic sound. [Beauchamp 1982] used the brightness to match the FM parameters, while [Horner *et al.* 1993] used genetic algorithms for the same task. This doesn’t make the underlying parameters more intuitive, however. [Ystad *et al.* 1996] matched additive parameters to a waveguide model parameter, which permits a greater intuitive understanding through physical parameters.

[Moorer 1976] introduced the discrete summation formulas, which are here called the brightness creation function. The easy calculation and recreation of brightness with these formulas, presented in this work, have not been found in the literature.

The chapter starts with a definition of the spectral attributes and the calculation formulas for the parameters used in the model in section 7.2. A brightness creation function in the additive and the time domain is presented in paragraph 7.2.2. The spectral envelope model is presented in section 7.3, along with the recreation of the spectral envelope in paragraph 7.3.5. The time-varying spectral envelope is presented in section 7.4. An initial study of a formant model is presented in section 7.5, and finally there is a conclusion.

7.2. Analysis of Perceptive Attributes

The spectral envelope is here defined as the maximum amplitudes of the harmonic additive parameters. The spectral envelope is an important attribute of the timbre of a sound [McAdams *et al.* 1995].

Figure 1 shows the spectral envelope for four typical musical sounds. It has, as can be seen, some features visible to the eye, such as the slope of the envelope and the irregularity of the spectrum. It also has a noticeable noise floor for some of the sounds, but this influences neither the analysis, nor the perception of the sound.

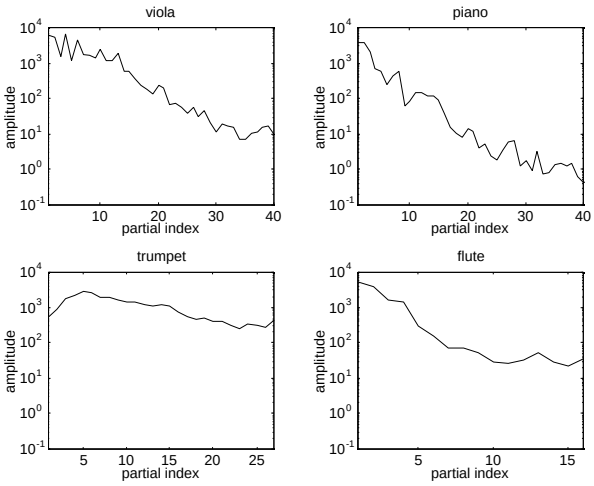


Figure 7.1. Spectral Envelope for the viola, the piano, the trumpet and the flute.

The trumpet has a strong resonance located around the fifth partial. It is shown in this work that the spectral envelope model presented here can model low partial resonances (formants).

The parameters of the spectral envelope model are found by analyzing the spectral envelope. The additive parameters are found using the linear time/frequency (LTF) analysis method presented in Chapter 4. Only the quasi-harmonic partials are saved. The amplitudes and frequencies of the analysis are $a_{k,t}$ and $f_{k,t}$, where *k* is the partial index and *t* is the time index. When the time index is omitted, the spectrum is supposed to be static. The static spectral envelope is here calculated as the maximum amplitude of each quasi-harmonic partial.

7.2.1. Brightness

Brightness is calculated as the spectral centroid, [Beauchamp 1982], which is correlated with the subjective quality brightness [McAdams *et al.* 1995]. The brightness is calculated as

$$brightness = \left(\sum_{k=1}^{N} k a_k\right) / \sum_{k=1}^{N} a_k \qquad \textbf{(7.1)}$$

A closely related attribute is sharpness [von Bismarck 1974b], which, like the brightness, correlates with the perception of brightness. If the partial multiplication k is replaced with the frequency of the partial, the brightness is expressed in Hertz. For harmonic sounds, this is equivalent to multiplying the partial index brightness with the fundamental. The partial index brightness is used in the rest of this work, if nothing else is stated. Other calculations of brightness can be done with the square amplitudes, with the log amplitudes, with real frequencies, instead of overtone index, as stated above, or with bark scale [Sekey *et al.* 1984] frequencies.

A good function to create additive parameters with a given brightness is,

$$a_k = B^{-k} \qquad \textbf{(7.2)}$$

The brightness of equation (7.2) has a simple expression, if the number of partials is set to infinity,

$$brightness = \frac{\sum k \;\; B^{-k}}{\sum B^{-k}} = \frac{B}{B-1} \qquad \textbf{(7.3)}$$

Brightness is thus infinity when B is 1 and decreasing when B is increasing. The value B is easy to calculate, if a given brightness T_b is researched,

$$B = \frac{T_b}{T_b - 1} \qquad \textbf{(7.4)}$$

The amplitudes found with the equation (7.2) have been calculated for the 4 sounds in figure 7.1. Only the quasi-harmonic partials have been included, and the x-axis is the partial index. The resulting curves are shown in figure 7.2. The brightness is indicated with a '*' at the x-axis.

The synthetic spectral envelope, recreated with the brightness only, restitutes much of the sound, but brightness alone is generally not enough to model a sound.

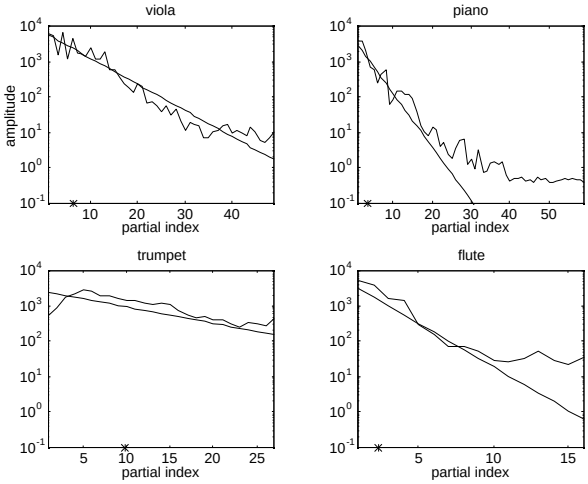


Figure 7.2. Spectral envelope for the viola, the piano, the trumpet and the flute with simple brightness matched curve. The brightness of each sound in marked with a '*'.

7.2.2. Time domain Brightness Function

Incidentally, the brightness function in equation (7.2) can be found in the time domain. This is done by setting the amplitudes as Diracs on the harmonic frequencies,

$$s_B(\omega) = B \sum_{k=1}^{\infty} B^{-k}(Dirac(\omega - k\omega_0) + Dirac(\omega + k\omega_0)) \qquad \textbf{(7.5)}$$

and taking the inverse fourier transform on that. The resulting time domain function is, after simplifications,

$$s_B(t) = \frac{1}{\pi} \frac{B\cos(\omega_0 t) - 1}{B^{-1} + B - 2\cos(\omega_0 t)} \qquad \textbf{(7.6)}$$

This function can easily be implemented with low processor cost. The time domain brightness function for a fixed brightness is shown in figure 7.3 (top), with the corresponding frequency magnitude (bottom). It is clear that this function has the characteristic linear frequency slope in the log amplitude domain. The resulting brightness for the equation (7.6) is given in equation (7.3). The value of B is found using equation (7.4). This is thus a very easy way of creating a time domain signal with a given brightness. The function given by the equation (7.6) is here called the brightness creation function (BCF). The BCF is equivalent to the discrete summation formulas presented in [Moorer 1976], which also gives the formulas for non-infinite summation, and for two sided spectra. [Moorer 1976] did not make the important connection between the discrete summation formulas and the brightness presented here.

The time domain signal for a varying brightness, going from 1 to 10 is shown in figure 7.4. The signal for brightness close to 1 is approximating a sinusoidal, as it should.

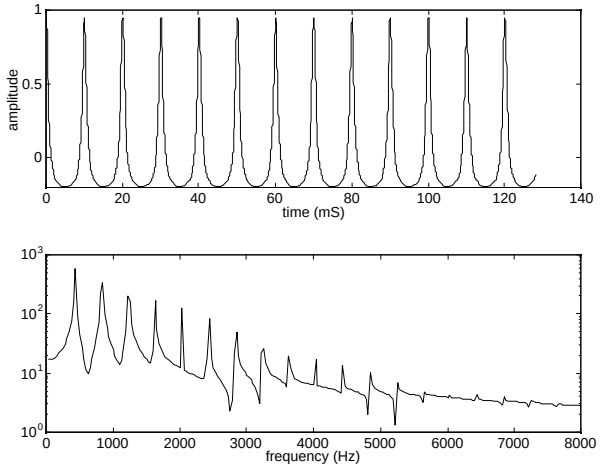


Figure 7.3. Time domain (top) and frequency domain brightness function. The partial index brightness is set to 3.0.

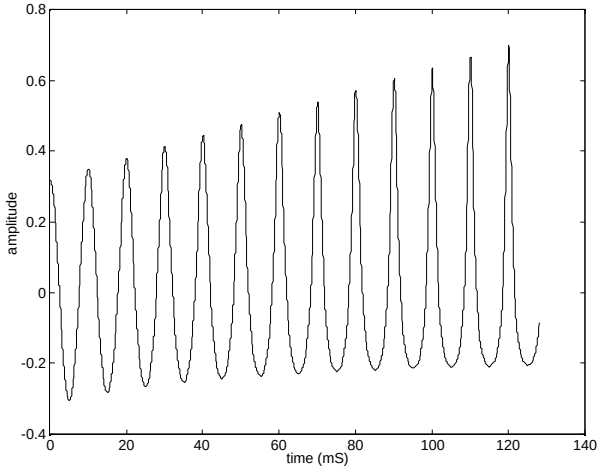


Figure 7.4. Time domain brightness function with variable brightness going from 1 to 10.

The time domain brightness function can of course be used to recreate a sound with a given (time varying) brightness, amplitude and frequency. For this, equation (7.6), multiplied with the amplitude, is used. The sound resulting from the BCF synthesis is of good quality, although rather clean, and missing the roughness and noise of the original sounds.

This function is considered here as very promising, if it is subsequently filtered to obtain the correct tristimulus and odd values, it could be a very cheap way of creating realistic musical sounds. The sum of several non-infinity summations could also improve the resynthesis quality. The BCF could be used as a source signal in both musical and speech sounds.

The BCF can give aliasing effects, if the brightness and/or the fundamental frequency are high. The aliasing are high partials above the nyquist frequency (sample rate/2) which are folded back into the audible spectrum. Although the aliasing might not be a problem in most cases, since the partial index brightness generally decreases with the fundamental frequency of a sound, if the BCF is to be put in use, the aliasing problem must be solved. No aliasing has been detected for the few sounds, which have been resynthesized using the BCF.

To verify whether a sound is causing aliasing effects, the amplitude of the partial at the nyquist frequency must be known. The amplitudes above this are all weaker than this amplitude.

The amplitude of the *k* partial is,

$$a_k = B^{-(k-1)} \tag{7.7}$$

and the partial index at the nyquist frequency is,

$$k_{nyquist} = \frac{samplerate/2}{f_0} \tag{7.8}$$

Aliasing does not occur, if the amplitude of the partial at the nyquist frequency is low,

$$a_{nyquist} = B^{-(k_{nyquist}-1)} < \varepsilon \tag{7.9}$$

When analyzing $a_{nyquist}$ for some instruments, it seems that the violin has a much higher value than other instruments (piano, clarinet, flute and soprano). The value of $a_{nyquist}$ is rather constant for an instrument, regardless of fundamental frequency. It is around 10^{-9} for the piano, 10^{-3} for the violin, 10^{-5} for the clarinet and the flute, and 10^{-7} for the soprano voice. This would mean that the amplitude at the hearing limit of a musical instrument is constant, regardless of fundamental frequency.

In conclusion, the BCF could probably be used for most musical sounds without disturbing aliasing effects.

Nevertheless, in some situations, the aliasing must be prevented. The spectrums from the BCF for 4 different brightness are shown in figure 7.5. The fundamental is 200 Hz, and the sample rate is 32 kHz. Aliasing occurs when the brightness value is above about 10.

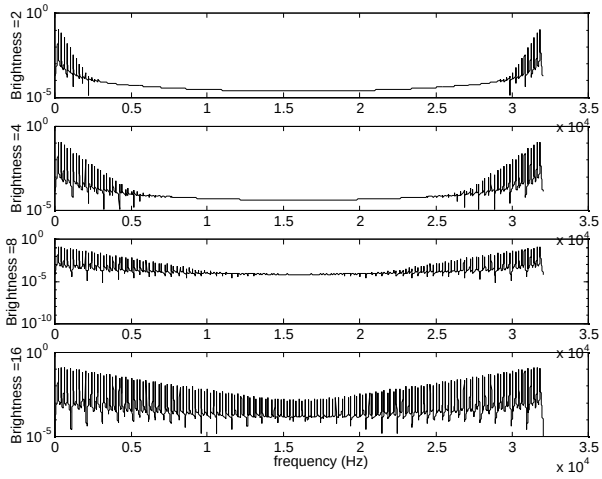


Figure 7.5. Resulting spectrum of the BCF for 4 signals with fundamental 200 Hz and sample rate 32 kHz. Brightness 2 (top), 4, 8 and 16 (bottom).

The aliasing could be prevented by taking the non-infinite sum of terms in equation (7.5) as proposed in [Moorer 1976]. The *B* term in equation (7.4) would then have to be recalculated.

Other techniques exist to do this, for instance, by bandlimiting the signal [Stilson *et al.* 1996]. This would prevent the BCF from having energy above the nyquist, regardless of brightness or frequency.

Although the time-domain BCF is very promising, for several reasons the timbre models used here are all modeled in the additive domain. No further use of this time-domain function is made in this work, but it could often replace the additive parameters with little or no loss of quality.

In conclusion, a time domain brightness creation function (BCF) has been presented. The spectrum of the BCF is linear in the log amplitude domain, and the brightness is easily calculated from, or given to the function. The combination of the BCF and dynamic filters could potentially create realistic musical instrument synthesis.

7.2.3. Tristimulus

The tristimulus values have been introduced in [Pollard *et al.* 1982] as a timbre equivalent to the color attributes in the vision. The tristimulus is used in [Pollard *et al.* 1982] to analyze the transient behavior of musical sounds. Other uses of the tristimulus includes the classification [Kostek *et al.* 1996], and the analysis of source spectrum of the flute [Ystad 1998]. The tristimulus are here defined as,

$$tristimulus1 = \frac{a_1}{\sum_{k=1}^{N} a_k} \quad \textbf{(7.10)}$$

$$tristimulus2 = \frac{a_2 + a_3 + a_4}{\sum_{k=1}^{N} a_k} \quad \textbf{(7.11)}$$

$$tristimulus3 = \frac{\sum_{k=5}^{N} a_k}{\sum_{k=1}^{N} a_k} \quad \textbf{(7.12)}$$

It is best plotted in a diagram where tristimulus 2 is a function of tristimulus 3. In such a diagram, the three corners of the low left triangle denote strong fundamental, strong mid-range, and strong high frequency partials. The tristimulus diagram can be seen in figure 7.6 along with the tristimulus for four musical instruments.

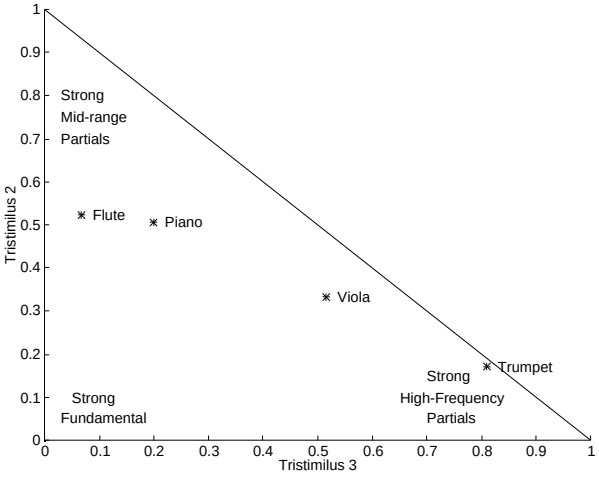


Figure 7.6. Tristimulus values for four sounds.

Notice that the sum of the three tristimulus equals 1. It is necessary only to use 2 out of the 3 tristimulus. Tristimulus 1 and 2 are saved.

7.2.4. Odd/Even Relation

The odd/even relation is well known from for instance, the lack of energy in the even partials of the clarinet [Benade *et al.* 1988]. To avoid too much correlation between the odd parameter and the tristimulus 1 parameter, the odd parameter is calculated from the third partial,

$$odd = \left(\sum_{k=2}^{N/2} a_{2k-1}\right) / \sum_{k=1}^{N} a_k \qquad \textbf{(7.13)}$$

$$even = \left(\sum_{k=1}^{N/2} a_{2k}\right) / \sum_{k=1}^{N} a_k \qquad \textbf{(7.14)}$$

Since tristimulus 1 + *odd* + *even* equals 1, it is necessary only to save one of the two relations. The odd parameter is saved.

7.2.5. Irregularity

Several studies have pointed at the importance of the irregularity of the spectrum [Krimphoff *et al.* 1994]. Irregularity is defined in [Krimphoff *et al.* 1994] as the sum of the amplitude minus the mean of the preceding, same and next amplitude,

$$irregularity = \sum_{k=2}^{N-1} \left| a_k - \frac{a_{k-1} + a_k + a_{k+1}}{3} \right| \qquad \textbf{(7.15)}$$

although in the log10 domain. In this paper, an alternative calculation of the irregularity is used, where the irregularity is the sum of the square of the difference in amplitude between adjoining partials,

$$irregularity = \left(\sum_{k=1}^{N} (a_k - a_{k+1})^2\right) / \sum_{k=1}^{N} a_k^2 \qquad \textbf{(7.16)}$$

and the N+1 partial is supposed to be zero. The irregularity value calculated in this way is most often, although not always, below 1. It is by definition always below 2.

Changing irregularity definitely changes the perceived timbre of the sound.

Irregularity changes the amplitude relations in the same tristimulus group. Since the tristimulus 2 value is large, this is where irregularity has the greatest influence.

The change of irregularity translates therefore principally into a change in the ratio between the second and the fourth partial amplitude. The third partial is fixed by the odd value.

The spectral envelope for 4 different values of the irregularity is shown in figure 7.7. The values of the brightness (5), the odd (0.3), and the tristimulus 1 (0.25) and 2 (0.5) are kept at the same value for all 4 plots. The perceived effect of the different values of the irregularity is rather big.

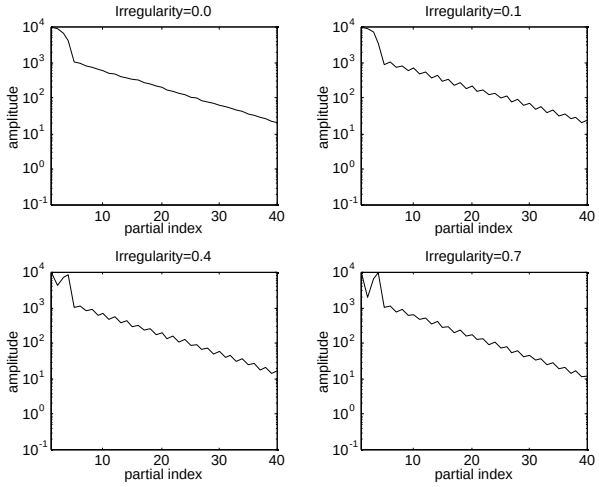


Figure 7.7. Spectral Envelope for four different irregularities, 0, 0.1, 0.4 and 0.7. Brightness=5, tristimulus 1=0.25, tristimulus 2=0.5, odd=0.3.

7.3. Spectral Envelope Model

This section presents a method of recreating N synthetic amplitudes, whose form, judged by the perceptual attributes, is similar to the original spectrum. The spectral envelope is modeled by the attributes presented in section 7.2; brightness, tristimulus1, tristimulus2, odd and irregularity. There are 5 attributes and N amplitudes. N is assumed to be greater than 5. There is thus an infinity of solutions. To limit the number of solutions,

the 5 to *N* upper harmonic components are set to an exponentially decreasing amplitude, and the 4 first harmonic components are then found.

The 4 perceptual attributes, omitting for the time the irregularity, are used,

$$T_b = brightness = (\sum_{k=1}^{N} k a_k) / \sum_{k=1}^{N} a_k \quad \textbf{(7.17)}$$

$$T_1 = tristimulus1 = a_1 / \sum_{k=1}^{N} a_k \quad \textbf{(7.18)}$$

$$T_2 = tristimulus2 = (a_2 + a_3 + a_4) / \sum_{k=1}^{N} a_k \quad \textbf{(7.19)}$$

$$T_o = odd = (\sum_{k=2}^{N/2} a_{2k-1}) / \sum_{k=1}^{N} a_k \quad \textbf{(7.20)}$$

7.3.1. The High Harmonic Components

In order to limit the number of solutions, the high harmonic components are set,

$$a_k = B^{-k}, k = 5,7,.., N \quad \textbf{(7.21)}$$

$$a_k = k_o B^{-k}, k = 6,8,.., N \quad \textbf{(7.22)}$$

where *B* is the brightness coefficient, and k_o is the odd coefficient. If the spectrum were to be defined by the formulas (7.21) and (7.22), although with the index *k* ranging from 1 and 2 respectively, *B* and k_o are calculated to be,

$$B = \frac{T_b}{T_b - 1} \quad \textbf{(7.23)}$$

$$k_o = (B + B^2 - 1)\frac{T_o}{1 - T_o} \quad \textbf{(7.24)}$$

The low harmonic components are then easily found as explained in 7.3.2, but, unfortunately, the resulting low amplitudes are sometimes negatives, as shown in 7.3.3. Another estimation of *B* and k_o is then necessary to find positive amplitudes.

7.3.2. The Low Harmonic Components

Given the equations (7.21) and (7.22), the equations (7.17) to (7.20) can now be rewritten in the matrix form,

$$\begin{bmatrix} T_1-1 & T_1 & T_1 & T_1 \\ T_2 & T_2-1 & T_2-1 & T_2-1 \\ T_b-1 & T_b-2 & T_b-3 & T_b-4 \\ T_o & T_o-1 & T_o & T_o-1 \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} \sum_{k=3}^{N} T_1 k_o B^{-(2k-1)} + \sum_{k=3}^{N} T_1 B^{-2k} \\ \sum_{k=3}^{N} T_2 k_o B^{-(2k-1)} + \sum_{k=3}^{N} T_2 B^{-2k} \\ \sum_{k=3}^{N} (T_b-(2k-1)) k_o B^{-(2k-1)} + \sum_{k=3}^{N} (T_b-(2k-1)) B^{-2k} \\ \sum_{k=3}^{N} T_o k_o B^{-(2k-1)} + \sum_{k=3}^{N} T_o B^{-(2k)} \end{bmatrix} \quad \textbf{(7.25)}$$

There are now 4 equations with 4 unknown, which are the amplitudes a_1 to a_4. The solutions for the amplitudes are, if N equals infinity, and after simplifications,

$$a_1 = \frac{(1+k_o B)T_1}{B^4(1-B^2)(T_1+T_2-1)} \quad \textbf{(7.26)}$$

$$a_2 = \frac{(1+k_o B)(4-3T_1-T_b-T_o+B^2(-6+5T_1+2T_2+T_b+T_o))}{2B^4(B^2-1)^2(T_1+T_2-1)} \quad \textbf{(7.27)}$$

$$a_3 = \frac{T_o+k_o B(1-T_1-T_b-T_o)}{B^4(B^2-1)(T_1+T_2-1)} \quad \textbf{(7.28)}$$

$$a_4 = \frac{-4+3T_1+2T_2+T_b-T_o+B^2(6-5T_1-4T_2-T_b+T_o)}{2B^4(B^2-1)^2(T_1+T_2-1)} + \frac{k_o B(-2+T_1+T_b-T_o+B^2(4-3T_1-2T_2-T_b+T_o))}{2B^4(B^2-1)^2(T_1+T_2-1)} \quad \textbf{(7.29)}$$

Unfortunately, the solutions sometimes have negative values, dependent on the initial values of B and k_o. The next paragraph will find B and k_o which will always give positive amplitudes.

7.3.3. Finding the Positive Range

The four low harmonic component amplitudes given by equations (7.26) to (7.29) always have a solution, but unfortunately, the solution sometimes gives negative values to one or more of the amplitudes. This problem can be solved by choosing the coefficients B and k_o so that the first four amplitudes are positive.

By analyzing the amplitude equations (7.26) to (7.29), it is found that a_1 is always positive, a_2 and a_3 are positive when B is smaller than lim_2 and lim_3 respectively, and a_4 is positive when B is greater than lim_4. The solutions for lim_2, lim_3 and lim_4 are found by

setting the corresponding amplitude to zero and isolating *B*. The limits are a function of T_1, T_2, T_o and T_b which are known, and k_o which is unknown.

The range of *B* for positive a_1 to a_4 is,

$$B_{range}(k_o) = [\lim_4, \min(\lim_2, \lim_3)] \quad \textbf{(7.30)}$$

The solutions to the limits lim_2 and lim_3 are,

$$\lim_2 = \frac{\sqrt{-4 + 3T_1 + T_b + T_o}}{\sqrt{-4 + 3T_1 + T_b + T_o} - \sqrt{-6 + 5T_1 + 2T_2 + T_b + T_o}} \quad \textbf{(7.31)}$$

$$\lim_3 = \frac{T_o}{k_o(T_1 + T_2 + T_o - 1) + T_o} \quad \textbf{(7.32)}$$

The solution for lim_4 is too long to be written here, but it is easily found using, for instance Mathematica [Wolfram 1996]. The range of *B* can of course be empty, in which case k_o is swept until a positive range is found. This gives a multitude of possible solutions, one of which must be chosen. The choice is made using the irregularity function, as shown in paragraph 7.3.4.

7.3.4. Finding Best Irregularity

The irregularity is the normalized square difference between the amplitudes of the partials. Of course, some irregularity originates from the brightness and the odd value, which gives a minimal value to the irregularity, but higher irregularity can be found by changing the values of *B* and k_o in equations (7.26) to (7.29).

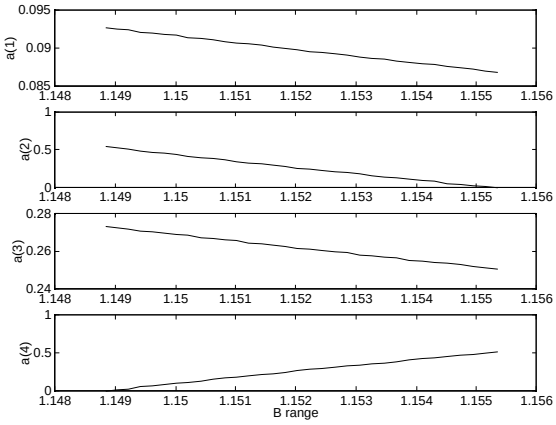


Figure 7.8. a(1) to a(4) in B_{range}.

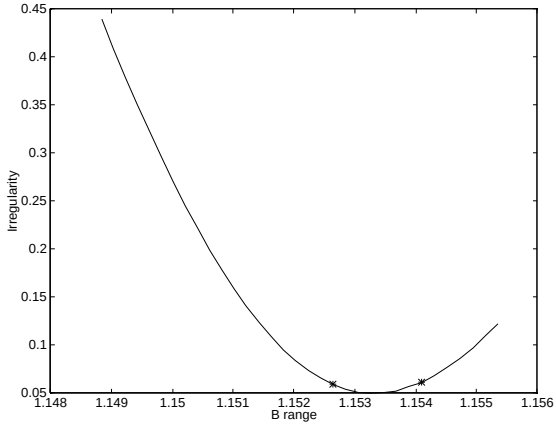


Figure 7.9. The irregularity) in B_{range}.

The irregularity is used to choose the values of *B* and k_o, but the different irregularity values also change the perceived effect of the sound.

With the results for the amplitudes in the preceding paragraphs, the irregularity can now be calculated for a given B and k_o. The goal is to find B and k_o that gives a wanted irregularity. Unfortunately, it is not possible to solve the irregularity equation, so B is swept over B_{range} for a number of k_o until the irregularity is correct. The irregularity form in B_{range} can be seen in figure 7.9. The correct irregularities are shown with '*'. There can be 0, 1 or 2 correct irregularities. If there is no solution, B_{range} is swept for a new k_o. If there is one solution, it is chosen. If there are 2 solutions, the solution closest to the middle is selected, since that would be the solution if there were only one correct irregularity. There are two solutions in figure 7.9, and the left solution is selected. k_o is swept, starting from 1, alternatively increasingly lower and higher than 1, until a solution is found.

7.3.5. Recreation of Spectral Envelope

With the B and k_o found, the synthetic amplitudes are now created, using the formulas (7.21), (7.22) and (7.26) to (7.29).

Since the spectral envelope parameters are not orthogonal, there are values that do not have a solution. These are generally the result of analysis of very low amplitudes, or the result of modifications of the spectral envelope parameters. If no solution is possible, the values of T_1, T_2, and T_o are slowly approached to a normalized value, and the spectral envelope creation is iterated until a solution is found. The default values of T_1, T_2, and T_o are the values these parameters would have in a clean BCF.

The synthetic spectral envelopes, which are here multiplied by the maximum of the original amplitude, can be seen in figure 7.10.

The sounds created from the spectral envelopes in figure 7.10 are very close to the original sounds from figure 7.1. Notice that the model of the spectral envelope is able to recreate the low formantic structure in the trumpet sound. The noise floor of the piano and the viola is of course not recreated, but this doesn't matter since it doesn't add to the sound quality, these partials being too low to be perceivable.

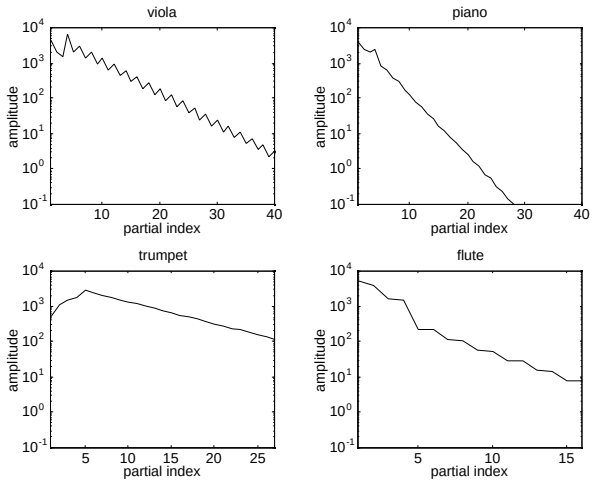


Figure 7.10. Synthetic spectral envelopes for the viola, the piano, the trumpet and the flute.

The resulting brightness of the modeled spectral envelope is always a little low, since there is a finite number of partials, and the model has been created with an infinity of partials. This could probably be adjusted, but the change in brightness is generally very small and it has not been judged to have any perceptual effect.

In conclusion, a spectral envelope model has been presented. It has a fixed parameter size. The parameters of the model permits a faithful recreation of the amplitudes from which the model parameters were found, including low formantic structures.

7.4. Time Varying Spectral Envelope

The spectral envelope model parameters can be calculated for the time-varying spectrum, and the synthetic time-varying amplitudes can be created from these parameters. This is a good test of the stability of the solution, and moreover, it permits listening to complete sounds, where a judgment can be made on, for instance the attack segment.

The time varying spectral envelope model parameters for the four test sounds can be seen in figure 7.11 for the viola, in figure 7.12 for the piano, in figure 7.13 for the trumpet and in figure 7.14 for the flute. The top left plot is the brightness, the top right plot is the tristimulus, the bottom left plot is the odd, and the bottom right plot is the irregularity. The tristimulus is plotted only for the times where the amplitude is above 10 percent of the maximum amplitude. There is no time axis for the tristimulus, where tristimulus 2 is plotted as a function of tristimulus 3, but the time can be followed from the start '+' to the end 'o'.

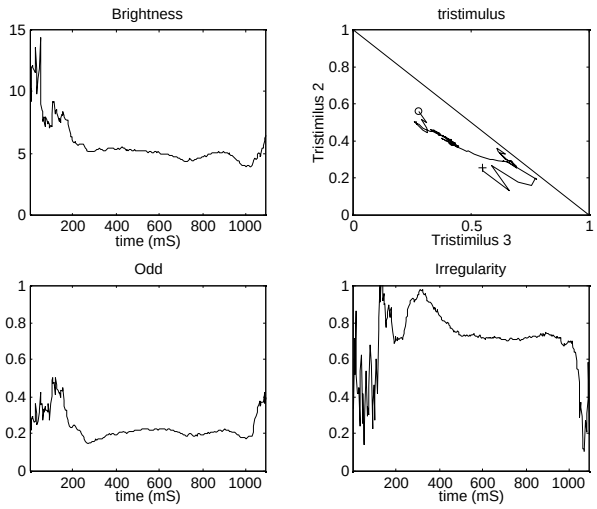


Figure 7.11. Time varying spectral envelope parameters for the viola.

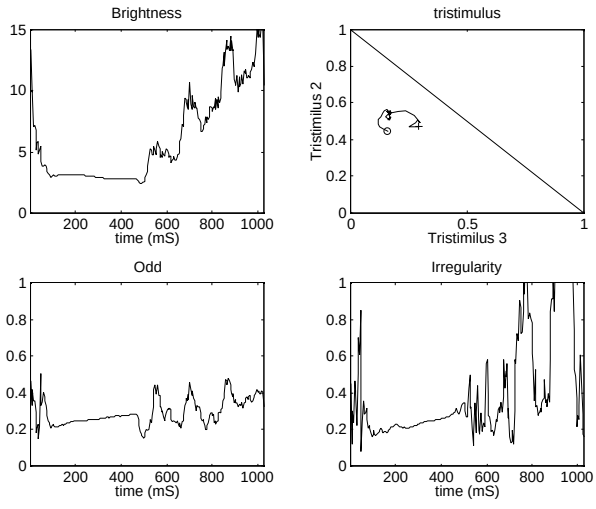


Figure 7.12. Time varying spectral envelope parameters for the piano.

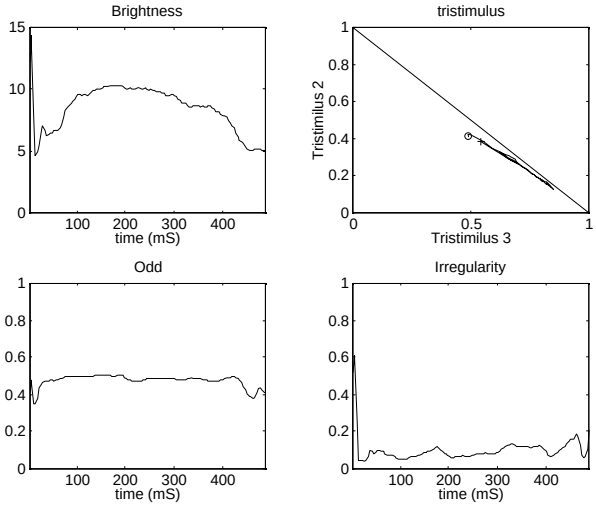


Figure 7.13. Time varying spectral envelope parameters for the trumpet.

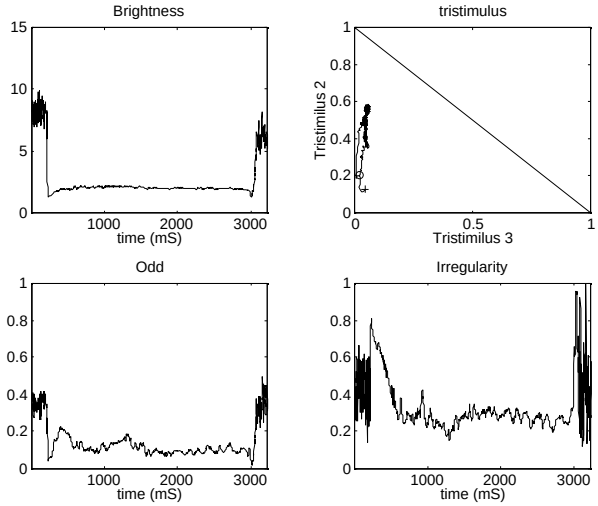


Figure 7.14. Time varying spectral envelope parameters for the flute.

The spectral envelope parameters seems rather stable in the sustain part of the sound.

The trumpet has much higher brightness in the middle of the sound than in the beginning and end of the sustain, even though the amplitude is rather stable throughout the sustain. The flute also has this behavior, although not as pronounced. The viola and the piano have falling brightness with time. These observations are made on the non-zero amplitude times, as observed in figure 7.15.

The viola has a lot of tristimulus variations, but most of this probably occurs in the attack. The trumpet has almost no tristimulus 1 and the flute has no tristimulus 3. The trumpet has a relatively high odd value, and the flute has a low odd value. The viola has a very high irregularity where the trumpet has a very low irregularity.

The recreated spectral envelope is normalized, and then multiplied by the time varying maximum amplitude of each sound, which can be seen in figure 7.15.

The amplitudes are rather smooth and stable in the sustain part of the sound for all instruments. This is not the amplitude of the fundamental or any one partial, but the maximum amplitude of all partials at each time segment. Nevertheless, many observations can be made from figure 7.15.

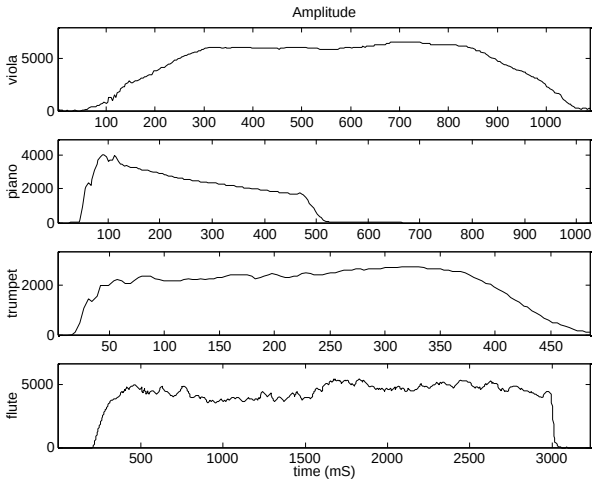


Figure 7.15. Time varying amplitude of the four test sounds. Viola (top), piano, trumpet and flute (bottom).

The viola has a relative slow attack and release, the piano has a rather fast attack and a decay slope. The trumpet attack is faster than the release, whereas the flute attack is much slower than the release. The flute has a pronounced irregularity (shimmer) on the amplitudes. This probably originates from the additive blowing noise of the flute. The additive noise is also present in the spectral envelope model parameters for the flute.

The recreated additive parameters of the spectral model parameters created from four sounds are shown in figure 7.16. The spectral model parameters have been calculated and the spectral envelope has been recreated for each time frame. The frequencies have been modeled using a simple model with the fundamental frequency and the inharmonicity for each time frame. More details on the frequency model can be found in Chapter 3.

The resynthesized sounds keep the realism of the original sounds, and are generally very hard to distinguish from the originals, although the noise is not modeled perfectly with this model. Including the tristimulus, the odd and the irregularity certainly improves the sound quality from the quality of the sounds created with the BCF in paragraph 7.2.2.

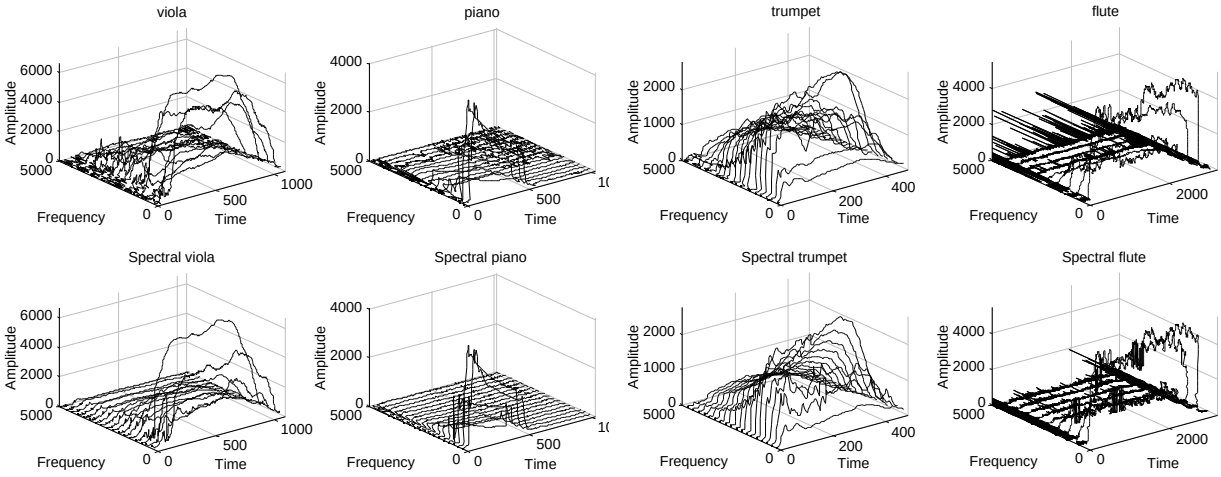


Figure 7.16. Original (top) and spectral envelope model (bottom) recreated additive parameters for 4 sounds, viola, piano, trumpet and flute.

The parameters of the spectral envelope model seem to recreate a stable spectral envelope for all the time frames. The sound quality of the spectral model resynthesis with frequencies using a simple model is significantly better than the sounds using static frequencies, but still not as good as the sounds recreated using the original frequencies.

7.5. Formants

The formants are resonant frequencies in the spectral envelope. The relative frequency and amplitude of the formants of speech defines which vowel is being pronounced

[Rabinet *et al.* 1978], [Klatt 1980]. Although few musical instruments have strong formants, a preliminary study of a formant model has been made, in the hope of being able to model the singing voice. The search for formants is made on the difference between the original spectral envelope and the recreated spectral envelope. The formants are supposed to be positioned above the 5th partial, and they can be both positive and negative. When a large error is found above the fifth partial, the error is modeled by a gaussian, and the gaussian is removed from the error signal. This is repeated until no more formants are found. The formants are now added to the modeled spectral envelope, and the process of finding formants is repeated.

Only formants wider than a threshold are saved. This decreases the chance of finding noise, or other spectral irregularities.

The formants are found by looking at the error signal,

$$e_k = \hat{a}_k - a_k \qquad \textbf{(7.33)}$$

Where $\hat{a}$ is the synthetic spectral envelope and a is the original spectral envelope. The maximum absolute error is now found, and modeled by a gaussian, which is defined by its amplitude, position and standard deviation,

$$t_g = a_g \; e^{\frac{-(k-k_0)^2}{2\sigma^2}} \qquad \textbf{(7.34)}$$

The amplitude of the gaussian is set to the error at position k_0, and the standard deviation is found by taking the mean of the standard deviations calculated on the left and the right of the maximum amplitude error. The gaussian is now subtracted from the error, and the new maximum error is found and modeled. This is repeated until the maximum error found is below a threshold. The formants are now defined by a sum of a few gaussians,

$$formants = \sum t_g(a_g, k_0, \sigma) \qquad \textbf{(7.35)}$$

and the new spectral envelope is

$$\hat{a}_f = \hat{a} + formants \qquad \textbf{(7.36)}$$

This creates a new spectral envelope including the strong formant regions. Although this method definitely reduces the error of the spectral envelope from the spectral model including the formants, there is no guarantee that the formants model really model formants, and not irregularities in the spectral envelope.

There is a higher probability of finding real formants if the sigma threshold of the gaussian is made bigger. In that way, only large formants are allowed. This could make the analysis miss peaked formants, though. The gaussian seem to match the error well in the few examples which have been tested, but no formal study of the validity of the gaussian form have been made.

The process of finding the formants is illustrated in figure 7.17 (top).

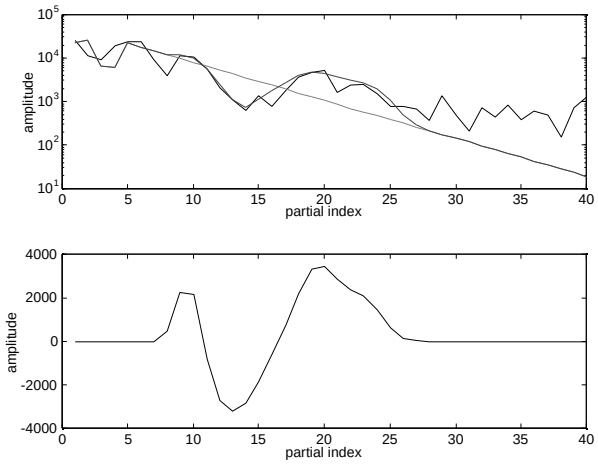


Figure 7.17. Illustrations of the formant search using a low 'a' sound. Top plot is original (solid), synthetic (dash-dotted), and with formants (dotted) spectral envelope, bottom plot is formants only.

The solid line is the original spectral envelope, the dashdotted line is the spectral model envelope and the dashed line is the spectral model and formant contribution spectral envelope. The bottom plot shows the formant contribution.

The formant analysis correctly finds two positive formants at 15 and 25, and a negative formant at around 10, although doubt could be expressed whether the positive formants should be stronger, and the non-formantic spectral envelope should be weaker above the 6th partial. The recreated spectral envelope is much closer to the original spectral envelope.

In conclusion, a formant analysis method has been described. It models the formants as a sum of gaussians in the linear amplitude domain. The addition of the formant model decreases the error of the spectral model, but the formants found by the formant analysis does not always correspond to the real formants. The formant model is not used in the rest of this work.

7.6. Conclusion

This work presents a new spectral envelope model. It models the spectral envelope with a few perceptually important attributes, but nevertheless, the visual shape of the envelope is often preserved. The parameters of the spectral model have been found in the literature of auditory perception, and they are brightness, the odd/even relation, tristimulus, and irregularity. The most important parameter is brightness, and functions for creating a signal with a given brightness have been found for the additive domain and for the time domain.

The combination of these parameters models a source spectrum, including formantic structure in the low partials. Most musical instruments can be modeled with this model, but unfortunately not the strong formantic structure of the human voice. An initial study of a formant model has therefore been performed. The formant model presented here reduces the error of the spectral model, but there is no guarantee it really models the formants and not other irregularities of the spectral envelope. The spectral envelope model with formants is not used in the rest of this document.

By analyzing a time-varying spectral envelope, a good restitution of harmonic sounds can be made. This indicates that this spectral envelope model is stable and well chosen.

Chapter Eight

8. Minimal Description Attributes

In this chapter, the parameters found in the HLA model presented in chapter 6 are further modeled by the partial evolution. This is done to extract the smallest number of parameters necessary to define a sound from the HLA model. Some of these parameters, such as the fundamental frequency, have an immediate perceptual value. Amplitude is modeled using the spectral envelope model in the preceding chapter, but most parameters keep the same unit as in the HLA model.

The Minimal Description Attributes (MDA) model generally has two values for each attribute, a fundamental value, and a partial evolution value. The fundamental value is useful when the value of an attribute is to be visualized, but the full MDA parameter set is necessary if a sound is to be resynthesized.

8.1. Introduction

The MDA model is an attempt to distill the minimum number of parameters necessary to characterize the identity and quality of an instrument. In order to do this, the high level

attributes calculated in the preceding chapter are used, but instead of keeping one parameter for each attribute and partial, a model of the curve along the partial axis for each attribute is found and modeled using few parameters.

The MDA model is created by curve fitting [Lancaster *et al.* 1986] the data of the HLA model to a simple curve. The simple curve should either have physical relevance [Fletcher *et al.* 1991] or minimize the perceptual error. The spectral envelope is modeled, as explained in Chapter 7, by minimizing the perceptual error using parameters correlated with perception [McAdams *et al.* 1995]. The frequencies are modeled using frequencies that corresponds to the frequencies of the quasi-harmonic partials of a stiff string [Fletcher 1964] or the frequencies of the impulse response of the flute [Ystad *et al.* 1996].

Not much literature involving the model of envelope or noise parameters as a function of partial index has been found. [Charbonneau 1981] models the attack and release times using a fourth order polynomial. [Ando *et al.* 1993] analyze the shimmer and jitter standard deviation, and plot it as a function of harmonic index, but they do not offer a model of the harmonic evolution.

The MDA model is kept as simple as possible. The amplitudes are modeled using the algorithms developed in Chapter 7, the frequencies are modeled using a simple stretched harmonics model, and the other parameters, including the envelope and the noise attributes, are modeled using a simple exponential curve. The exponential curve has been chosen from a selection of linear, polynomial and other curves by performing informal listening tests.

In addition to the parameters describing the amplitudes, frequencies and the exponential curve for all other attributes, an error term is also calculated for each attribute. This error term can be used, in theory, to recreate several variations of the same performance, in the same manner that an instrumentalist never sounds exactly the same each time he or she plays a note. This corresponds to the variants in [Risset *et al.* 1982].

This chapter starts with a definition of the frequency model in section 8.2. The spectral envelope model is discussed in section 8.3. The generic model using an exponential curve is presented in section 8.4, and the error term is discussed in section 8.5. The analysis from the HLA model is detailed in section 8.6 and the recreation of HLA models is explained in section 8.7. The sound quality of the MDA model is discussed in section 8.8, and the chapter ends with a conclusion.

8.2. Frequency Model

The frequencies of the partials are important to the perceived timbre of the sounds. A simple model of the frequencies is used here, which can model quasi-harmonic sounds with stretched or compressed harmonic frequencies. The frequencies are characterized by 2 parameters, the fundamental frequency and the inharmonicity, which is a measure of how much the higher partials are 'stretched' above or 'compressed' below the ideal harmonic frequency.

The frequency model is based on a model for the frequencies of a stiff string [Fletcher 1964]. The frequency of the partial *k* of a stiff string is,

$$f_k = k f_0 \sqrt{1 + \beta k^2} \qquad \textbf{(8.1)}$$

where f_0 is the fundamental frequency and β is the inharmonicity. The values of f_0 and β are found using a nonlinear least-squares curve fit [Moré 1977]. The same model is used when analyzing the frequencies of the impulse response of the flute [Ystad *et al.* 1996], although β is here negative.

The frequencies, divided by the partial index, for 4 musical instrument sounds, can be seen in figure 8.1, along with the MDA model frequencies.

The inharmonicity for the piano is easy to see. Notice the y-axis scale for the piano, since what is shown is the partial frequency divided by the partial index, the difference between two tones in the 40th partial is 320 Hz, 60 Hz more than the fundamental, which is about 260 Hz.

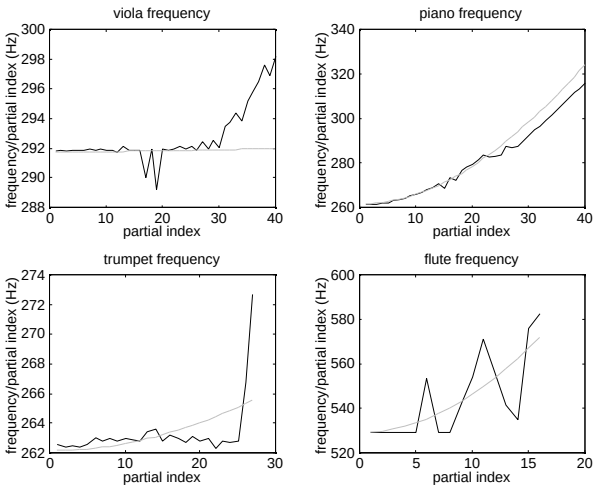


Figure 8.1. Analyzed frequency (solid), and MDA model frequency (dotted) for 4 instrument sounds, viola, piano, trumpet and flute.

The weak high order partial frequencies can be misjudged due to the presence of noise, and they are not used in the curve fit. The estimated frequencies can be seen in the dotted line.

The presence of inharmonicity in the piano certainly adds a flavor to the sound, and it is necessary to use the inharmonicity model, instead of assuming clean harmonic frequencies.

The viola seems perfectly harmonic, the piano has stretched partial frequencies, the trumpet also has almost perfectly harmonic frequencies, whereas the frequencies of the upper partials of the flute have been misjudged due to noise, and the flute inharmonicity value is estimated to be non-zero. The values of the inharmonicity of the four sounds are $1.2\ 10^{-06}$, $3.4\ 10^{-04}$, $3.6\ 10^{-05}$ and $6.6\ 10^{-04}$. The flute has erroneously a higher inharmonicity value than the piano.

8.3. Amplitude Model

The amplitudes are described by the spectral envelope model introduced in Chapter 7. The sounds are assumed to lack a formantic structure, or other resonant behavior, although provisions for formants have been made. The attributes describing the amplitudes are brightness [Beauchamp 1982], tristimulus [Pollard *et al.* 1982], the odd/even relation [Fletcher *et al.* 1991] and irregularity [Krimphoff *et al.* 1994]. See [McAdams *et al.* 1995] for a review of these and other timbre attributes. The formulas for the spectral envelope attribute calculations are,

$$T_b = brightness = \left(\sum_{k=1}^{N} k a_k\right) / \sum_{k=1}^{N} a_k \qquad \textbf{(8.2)}$$

$$T_1 = tristimulus1 = a_1 / \sum_{k=1}^{N} a_k \qquad \textbf{(8.3)}$$

$$T_2 = tristimulus2 = (a_2 + a_3 + a_4) / \sum_{k=1}^{N} a_k \qquad \textbf{(8.4)}$$

$$T_o = odd = \left(\sum_{2}^{N/2} a_{2k-1}\right) / \sum_{k=1}^{N} a_k \qquad \textbf{(8.5)}$$

$$irregularity = \left(\sum_{1}^{N} (a_k - a_{k+1})^2\right) / \sum_{k=1}^{N} a_k^2 \qquad \textbf{(8.6)}$$

For the recreation of the amplitudes from the spectral envelope attributes, see the spectral envelope model in Chapter 7. The recreation creates amplitudes, which are exponentially decaying, combined with an odd/even relation, above the 5th partial, whereas the first 5 partials have an individual shape. The recreations usually keep the shape of the spectral envelope, and more important, since they are derived from perceptual research, the perceptual quality of the sound.

The original spectral envelopes of 4 musical instruments are shown in figure 8.2, along with the MDA model spectral envelope (dotted). It is interesting to see that the different spectral parameters permit a good resynthesis of a formant, if it is below the 5th harmonic, as seen in the trumpet example.

The perceptive spectral envelope model parameters keep the original values for all sounds.

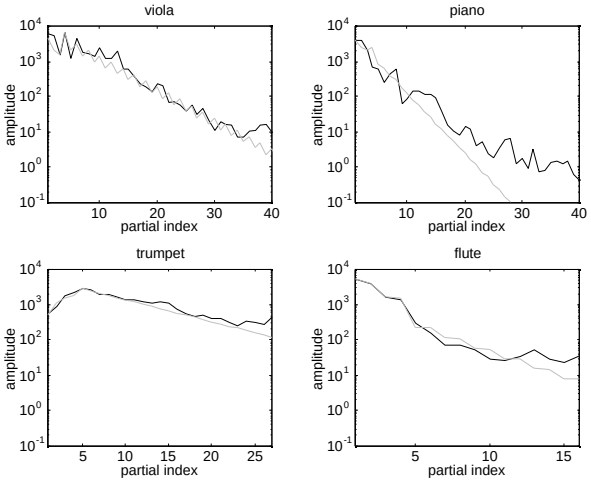


Figure 8.2. Spectral envelope four 4 musical instruments, with the MDA model spectral envelope (dotted).

8.4. Generic Parameter Model

The rest of the parameters in the HLA model are modeled by a simple exponential curve with 2 parameters,

$$c_k = v_0 * e^{v_1 k} \qquad \textbf{(8.7)}$$

where k is the partial index, v_0 and v_1 are the parameters of the MDA model curve c_k.

To estimate the parameters v_0 and v_1, an initial estimation is first found by linear least square curve fit [Schwarz 1989] in the log domain. This initial estimation is then used in a non-linear least square curve fit [Moré 1977] with the original measured values.

The estimation of the parameters is improved by using only the strong partials of the sound, as explained in paragraph 8.4.1.3.

8.4.1. Envelope Parameters

The envelope is the time-varying amplitude of each partial. The envelopes are here normalized between zero and one. The maximum amplitude of each partial is stored in the spectral envelope.

The parameters of the envelope model are the envelope times, the envelope relative amplitudes (percents), and the envelope segments curve forms.

8.4.1.1 Envelope Times

The parameters that are estimated here are the start time, the attack time, the sustain time, the release time, and the total length. Furthermore the start of attack, end of attack, start of release and end of release percents of the maximum amplitude are modeled along with the curve form for the 5 segments.

The times are modeled as the relative times between segments for the attack and release, and as the absolute times for the start, the sustain and the end. The attack and release times for the 4 sounds are plotted in figure 8.3 along with the MDA model times (dotted). The HLA parameters are rather noisy, so the exponential model is rather unfounded in some situations, but the attribute values seem to be respected in general.

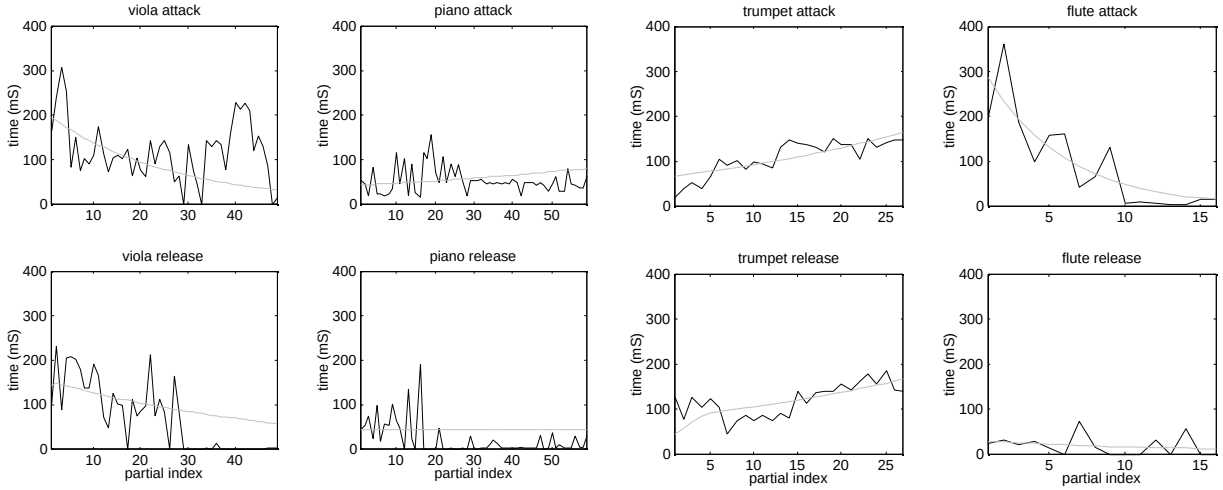


Figure 8.3. Attack and release times for the 4 sounds, with the MDA model envelope times (dotted). Attack (top) and release (bottom).

The envelope times are generally well modeled with the exponential curve. The flute attack is especially well modeled. However, noise and bad analysis often perturb the envelope time values; this is the case for the piano attack, for instance. All in all, the envelope times are well modeled by the simple exponential curve. Notice the atypical behavior of the trumpet attack and release, where the high partials are slower than the low partials. The release times deviate in the low partials of the trumpet, since the total time of all five segments otherwise would have been greater than the end time.

8.4.1.2 Envelope Percents and Curve Forms

The percents are the relative amplitudes of the partials at the split points. There are 4 percents, for the start of attack, the end of attack, the start of release and the end of release. The percents multiplied by the spectral envelope value for the same partial yields the split point amplitude.

The curve form is the shape of the segments. There are five curve forms, for the start, the attack, the sustain, the release and the end segments.

The end of attack and start of release relative amplitudes for the 4 sounds are plotted in figure 8.4 with the MDA model parameters (dotted), and the attack and release curve forms are plotted in figure 8.5 with the MDA model parameters (dotted). The curve form values are set to a default value (1) if the curve is too short. This is visible in the upper partials of the piano.

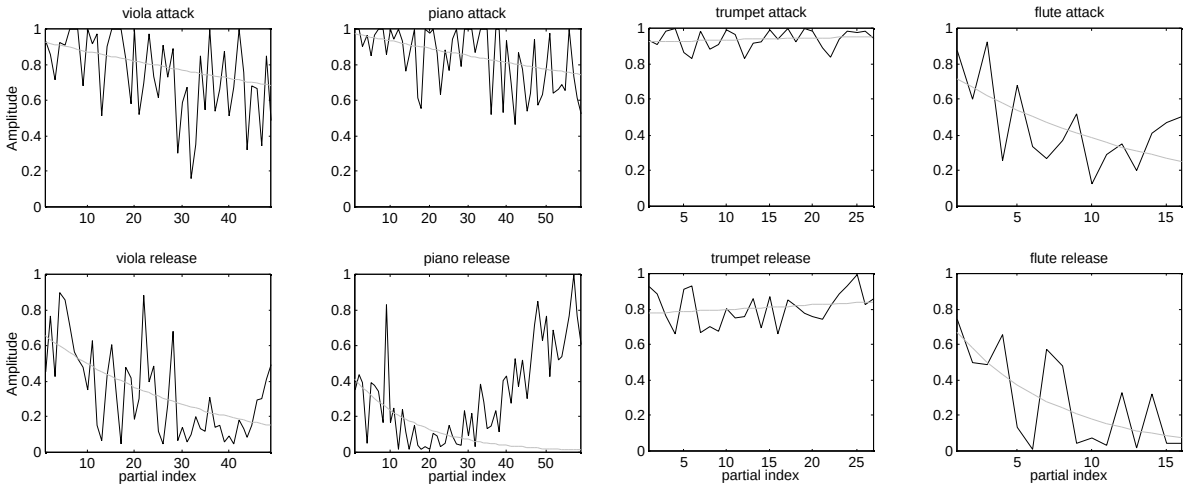


Figure 8.4. End of attack (top) and start of release (bottom) percents for the 4 sounds with the MDA model parameters (dotted).

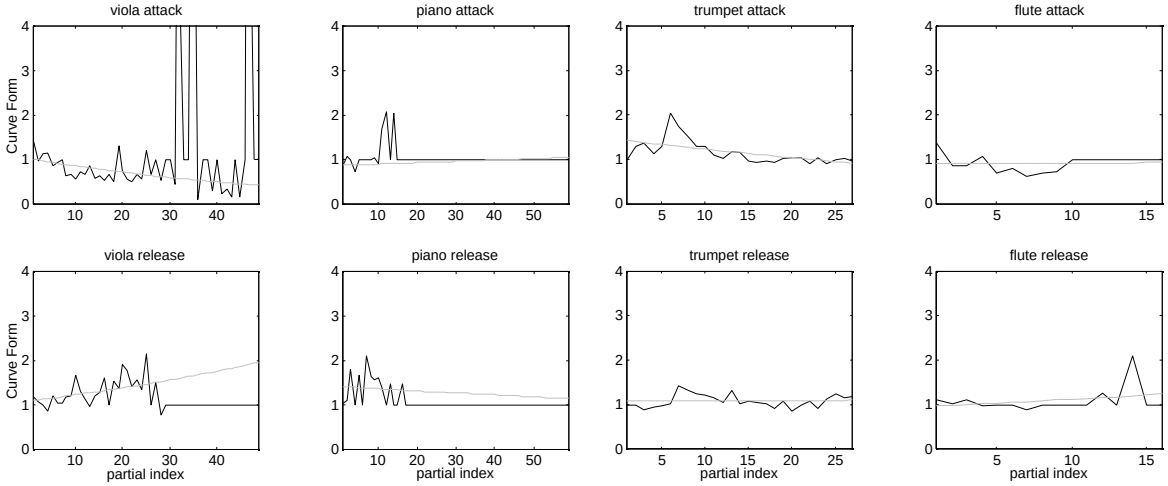


Figure 8.5. Attack and release curve form for the 4 sounds with the MDA model parameters (dotted).

The percents seem to fit the exponential model well. The improvement explained in 8.4.1.3 is very visible in the piano release relative amplitudes.

The viola, piano and flute percents are falling with frequency, whereas the trumpet release is rising with frequency. The trumpet attack percents seem constant. The percents seem rather correlated with the attack and release times in figure 8.3.

The double curve in the release percents for the piano is probably explained by noise in the higher partials, which could increase the percent values.

The exponential model generally seems to fit the curve forms very well, although mostly because the curve form values are relatively constant. The viola curve form values seems rather exponential, the piano values are not very reliable because of the short piano attack and the relatively important transient behavior.

The trumpet and the flute curve form values have a shape which seems reliable and which is not modeled by the exponential curve. Nevertheless, the deviations are small and this has not been found perceptually important.

8.4.1.3 Weak Partials

The weak upper partials often disturb the estimation of the parameters of the exponential model. For this reason, they are removed from the data before the estimation.

One important example is found in the release percents for the piano.

The upper partials of the piano are very weak, and more sensitive to noise and bad analysis.

The piano release percents are shown in figure 8.6 with the recreated percents made with all partials (dashdotted), and with only the first 32 strong partials (dotted).

The curve forms are obviously very different, and the dotted curve would be preferred, since the high partials, which are modeled better with the model using all partials, are relatively weak and inaudible.

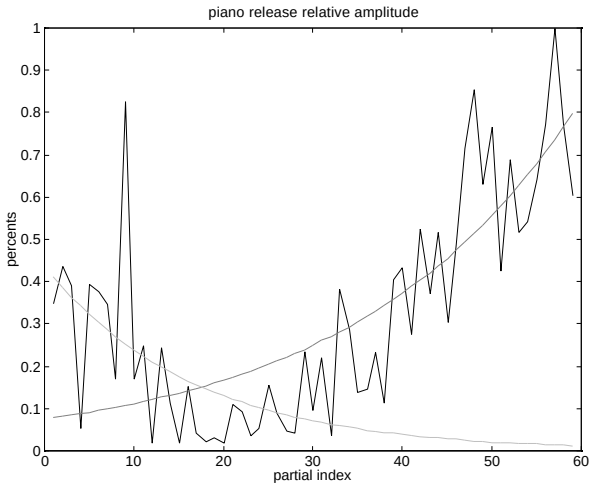


Figure 8.6. Piano release percents (solid) with all partials model(dashdotted) and 32 first partials model (dotted).

The modeling of all exponential curves are therefore made using only the partials whose amplitude is above a threshold relative to the maximum amplitude. This improvement is used in the rest of this work.

8.4.2. Noise Parameters

The noise on the frequency of the partials (jitter) and the noise on the amplitude of the partials (shimmer) are modeled in the attack, sustain and release segments by 2 parameters in each segment, one for the standard deviation and one for the single-tap recursive filter coefficient. Furthermore, the correlation between the fundamental and the other partials shimmer and jitter are modeled for the full length of each partial.

The shimmer parameters for the sustain part of the sound are plotted in figure 8.7, with the standard deviation (top), the filter coefficient (middle), and the correlation (bottom). Some of the partials are too short to permit an estimation of the noise parameters; these are therefore set to default values.

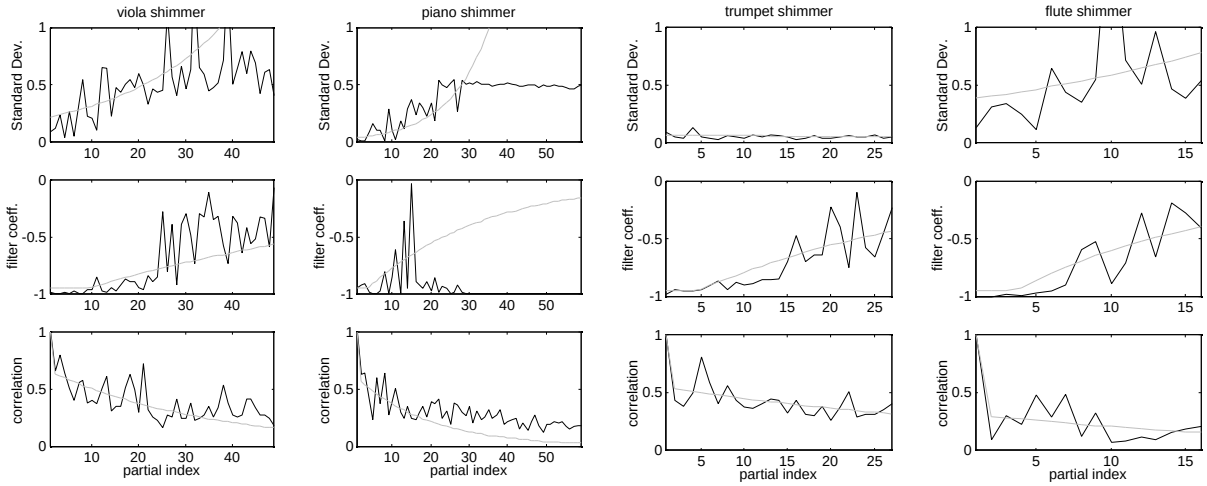


Figure 8.7. Sustain shimmer parameters for the 4 instruments with the MDA values (dotted). Standard deviation (top), filter coefficient (middle) and correlation (bottom).

The shimmer parameters don't seem to fit the exponential model very well, with the exception of the correlation. This is in part explained by noise on the high and weak partials, but still, it seems that the important noise parameters need another model, which fits the data better.

The jitter parameters for the sustain part of the four sounds are plotted in figure 8.8. The standard deviations are plotted on top, the filter coefficients are plotted in the middle and the correlations are plotted in the bottom. The jitter model has the same problem as the shimmer model. The data is not very exponential, and very noisy, so the recreated curve sometimes fits the noise of the curves more than the important partial values.

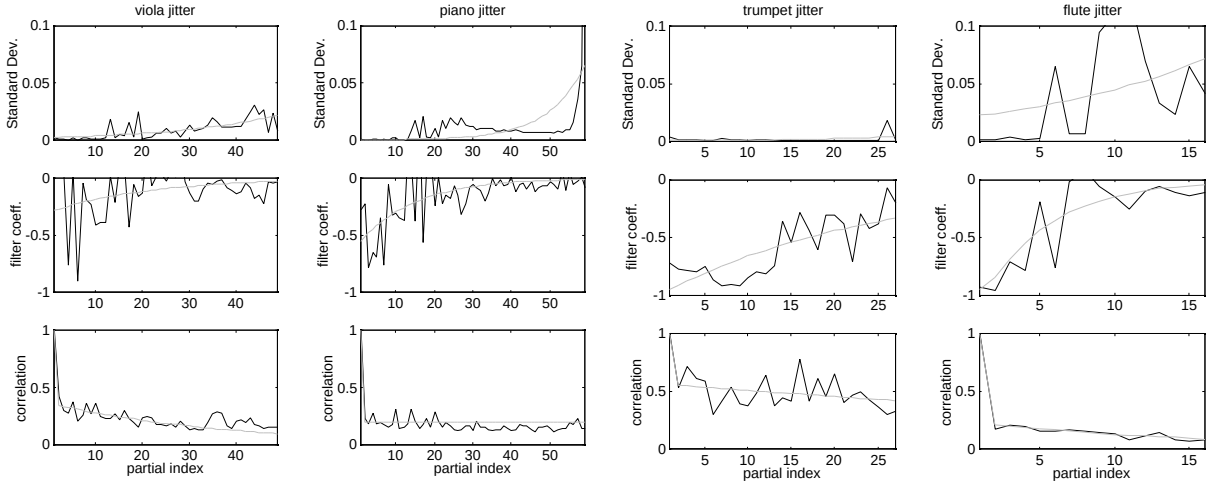


Figure 8.8. Sustain Jitter parameters for the 4 instruments with the MDA values (dotted). Standard deviation (top), filter coefficient (middle) and correlation (bottom).

8.4.3. Comments on the Noise Model

The noise parameters are very sensitive to the weight of the timbre attribute parameters. One important example is the sustain jitter standard deviation, as seen in figure 8.8. The lower partials obviously have very little jitter for all four instruments, and the higher order partials have an important jitter standard deviation. When modeling this with the exponential form, the lower partials get a too large jitter standard deviation, especially for the viola and the flute sounds. This completely changes the perception of the noise of these instruments; the relatively high-frequency noise is transformed into a low-frequency rumble.

An alternative to the exponential curve could be, for instance, a polynomial. A second order polynomial has been tested with good results. The exponential curve given in equation (8.7) is replaced with the second order polynomial given by,

$$c_k = v_0 + v_1 k + v_2 k^2 \qquad \textbf{(8.8)}$$

where k is the partial index. The parameters of the polynomial model are found using the linear least-squares fit [Schwarz 1989]

The standard deviation of the shimmer and the jitter of the flute are plotted in figure 8.9 for the exponential model (dotted) and the second order polynomial model (dashdotted). The lowest partials get less noise with the polynomial model, which generally seems to fit the data better. When listening to sounds recreated with the two models, the difference in sound quality is important. Where the exponential model has got a rumble quality to the sound, the second order polynomial noise quality is much closer to the original.

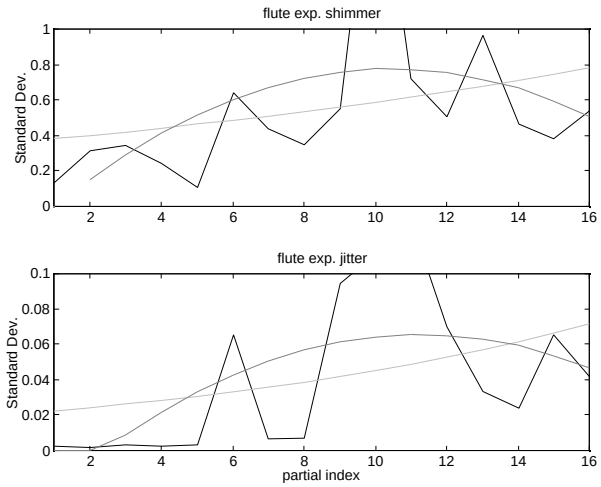


Figure 8.9. Flute shimmer (top) and jitter (bottom) with exp. model (dotted) and 2nd order polynomial model (dashdotted).

This is because the polynomial model restitutes the low jitter and shimmer of the strong low partials, whereas the exponential model gives the low partials too much jitter and shimmer.

The different models don't change much for the other three test sounds. Unfortunately, the polynomial noise model was introduced after the listening tests performed in Chapter 12 so no objective measure of its quality has been made. It seems, definitely, that the exponential model is not suited for the modeling of the important noise standard deviations. A second order polynomial performs better. Another option is to weight the HLA values before the curve fit.

8.5. Error Term Calculation

Although the curves above fit the data in the least-squares sense, they are by no means equivalents. The difference between the clean exponential curve and the data is assumed to be related to the execution of the sound, and the error can, if modeled properly, introduce new executions of the same sound, i.e. of the same instrument, player and style, in the same environment.

The error between the data curve and the exponential curve is supposed to be normal distributed, and it is modeled by the standard deviation. The error is furthermore divided into an odd and an even error, which have a separate mean value. This is done so that for instance bad analysis of the weak even partials of the clarinet will not introduce too much noise in the strong odd partials. The error is weighted by dividing by the partial index, and

recreated by multiplying by the partial index. This ensures that the normally strong lower partials don't get a high error from the weak, and error prone, higher partials.

Although it is possible to recreate a deviation from the exponential curve, which is similar to the error, it does not always give the same perceptual quality. One reason for this could be that the error is not random, but instead correlated, either between attributes, or between sounds from the same instrument. It could also be suspected that the error term is related to the model of the sounds, or the estimation of the model parameters.

One way of finding out is to look at the correlation between different error terms. The correlation of the different timbre attribute errors for 4 musical instrument sounds are analyzed here.

The error for the attribute *i* and the partial *k* is

$$e_k^i = (p_k^i - v_k^i)/k \qquad \textbf{(8.9)}$$

where p^i is the HLA timbre attribute *i* and v^i is the MDA modeled timbre attribute *i*. The error has *N*x*M* terms, *N* is the number of partials and *M* is the number of timbre attributes. The error *i* is said to be mostly correlated with another timbre attribute *j* when

$$j = \max(correlation(e^i, e^{i \neq j})) \qquad \textbf{(8.10)}$$

When analyzing the correlation of the errors, the different timbre classes are in general mostly correlated with another attribute from the same timbre class, for instance the sustain time is mostly correlated with the release time for 3 out of 4 instruments. There are exceptions: the release curve form error is mostly correlated with the release shimmer std, which leads to the conclusion that if the curve form is wrong, then the error is important. The start of attack percent is correlated with an attack error for 3 out of 4 instruments, which indicates again that if the envelope is wrong then the error is large. This is probably an analysis error, and not a feature of the sounds. The sustain and release shimmer std is mostly correlated with envelope attributes for 7 out of 8 correlations. The release jitter filter coefficient is mostly correlated with 4 envelope attributes.

In conclusion, it seems that the error is dependent more on the analysis, than on the actual quality of the sound. Further work is needed in order to use the error term successfully in the resynthesis of sounds.

8.6. Analysis from HLA Attributes

The HLA parameters presented in Chapter 6 are all that is needed to create the MDA parameters. The envelope times have been modified, so that the attack, sustain and release times are relative, and not absolute. The amplitudes and frequencies are analyzed by the methods in sections 8.2 and 8.3 and the other parameters are calculated by the methods exposed in section 8.4. Care must be taken when estimating the filter coefficient parameters, since the filter coefficients are negative.

8.7. Recreation of HLA Attributes

The recreation of the HLA parameters is straightforward, once some simple numerical limits are respected. All the parameters must be positive, with the exception of the filter coefficients, which lie between -1 and 0. The noise correlation values are always between zero and one, with the first value, which is not used in the curve fit, equal to one. The frequencies are calculated by the formula (8.1). The amplitudes are calculated by the method introduced in Chapter 7.

The complete HLA parameter set, recreated from the MDA parameters, without error, for the same four sounds as in the Chapter 6, is plotted in figure 8.10, figure 8.11, figure 8.12 and figure 8.13. The corresponding HLA parameters with error in the parameters are plotted in figure 8.14, figure 8.15, figure 8.16 and figure 8.17.

The spectral envelope (top left) and the frequencies (second from top left) have only one curve each. The envelope timing (third from top left) has four curves, the start of attack time ‘o’, the end of attack time ‘*’, the start of release time ‘x’, and the end of release ‘+’. The percents (fourth from top left) also have four curves with the same symbols, the curve forms (bottom left) have 5 curves, the start curve ‘+’ the attack curve ‘o’, the sustain curve ‘*’, the release curve ‘x’ and the end curve ‘.’. For the sake of clarity, the start and end curves are dotted. The shimmer std (top right), jitter std (second from top right), shimmer filter coefficients (third from top right) and), jitter filter coefficients (fourth from top right) have 3 curves, attack ‘o’, sustain ‘*’, and release ‘x’. The noise correlation is shown bottom right with the shimmer ‘o’ and the jitter ‘*’.

The spectral envelopes for the four sounds are quite different. The flute has no amplitude in the high partials, where the viola has relatively high amplitude for the 16 partial. The trumpet has a very visible formant region around the fifth partial.

Both the flute and the piano have stretched frequencies, but in the case of the flute it is because of noise on the weak upper partials, whereas the piano really has a stretched frequency curve.

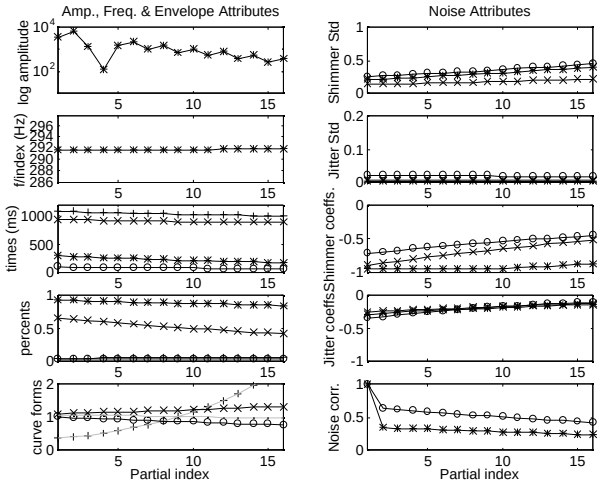


Figure 8.10. Recreated HLA parameters for the viola.

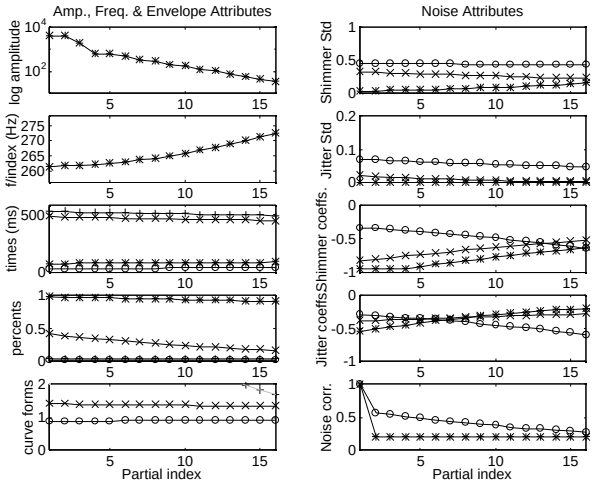


Figure 8.11. Recreated HLA parameters for the piano.

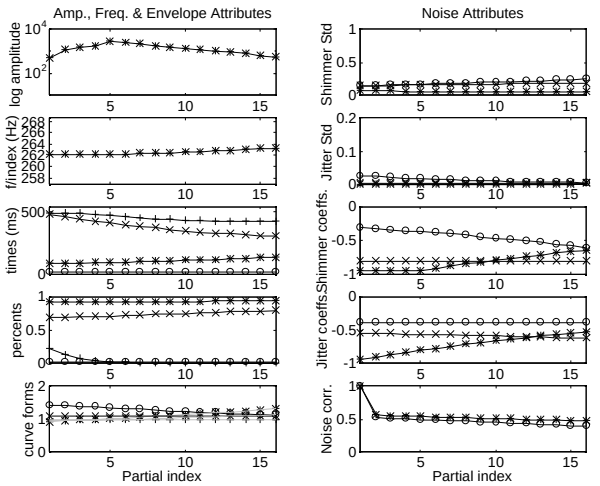


Figure 8.12. Recreated HLA parameters for the trumpet.

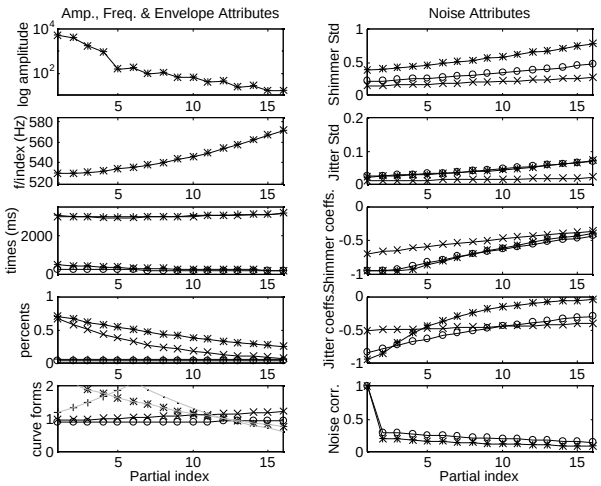


Figure 8.13. Recreated HLA parameters for the flute.

The piano has rather static envelope times for all partials. The trumpet envelope times have the typical shape, where the higher partials attack ends later and the release starts earlier. The flute envelope times are difficult to see, because of the relatively long sound, but the low partials attack seems longer than the high partials attack. The viola also has shorter upper partial attacks.

The piano percents are very low for the start of release split point. The start of attack and end of release percents are low for all sounds, whereas the end of attack is high for all

sounds except the flute, where both the end of attack and the start of release percents drop with frequency.

The curve form values are generally close to one for the attack of all sounds. The start and end values are not very important, since they model segments that are close to silent. The trumpet attack is exponential, as is the viola and piano releases.

The shimmer std values are much higher than the jitter std values, as could be expected. The trumpet has a relatively low shimmer std, and the flute has a relatively high shimmer std. The piano has very high shimmer and jitter std for the attack and release segments.

The jitter filter coefficients are generally higher than the shimmer filter coefficients. This can be explained by the simple amplitude model, which gives the shimmer a low frequency quality.

Correlation is generally lower for the jitter than for the shimmer. This could again be explained by the simple amplitude model. The frequency model is of course very good, since the frequencies are supposed to be static.

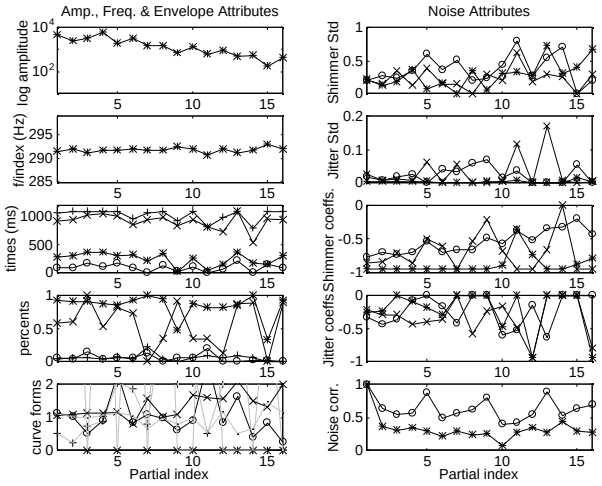


Figure 8.14. Recreated HLA parameters of the viola, with error term.

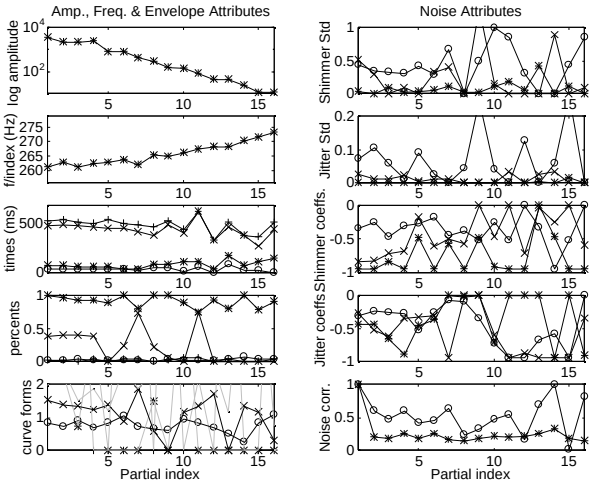


Figure 8.15 Recreated HLA parameters of the piano, with error term.

The MDA parameters with error term are often quite close visually to the corresponding HLA parameters, which can be seen in section 6.7 on page 79 in Chapter 6. The amount of noise is an indication of the success of the model parameter estimation, as shown in section 8.5. The spectral envelope and frequency curves seem relatively clean, and the envelope parameters are also generally rather clean, with the exception of the curve forms, although the important attack curve form curve is still visible.

There is more error on the noise parameters, with the exception of the correlation. The filter coefficients are very noisy.

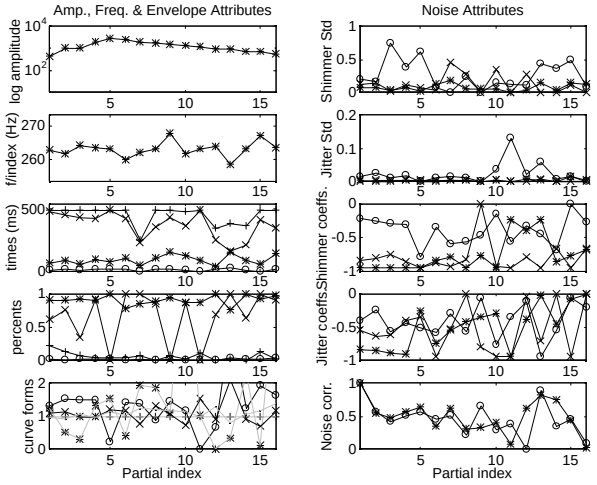


Figure 8.16 Recreated HLA parameters of the trumpet, with error term.

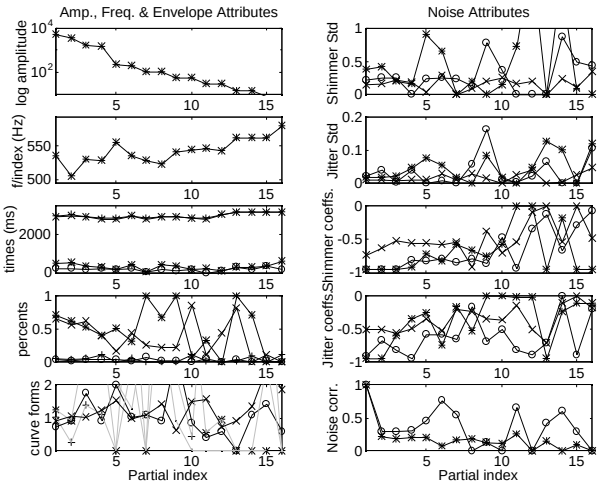


Figure 8.17 Recreated HLA parameters of the flute, with error term.

All MDA parameters have the same amount of noise, maybe with the exception of the trumpet, which seems cleaner.

8.8. Sound Synthesis from the MDA

The sound synthesis from the MDA is done through the HLA and the additive parameters as described in Chapter 6. In principle the MDAs create a sound with the same complexity as the other models. In figure 8.18 (viola), figure 8.19 (piano), figure 8.20 (trumpet) and figure 8.21 (flute) the additive parameters from the original analysis (left), the recreated additive parameters from the MDA parameters without error term (middle) and with error term (right) are shown.

Visually, the viola, piano and trumpet seems to have kept the shape of the parameters, whereas the flute is very distorted by noise. This translates into a greater impairment of the sound.

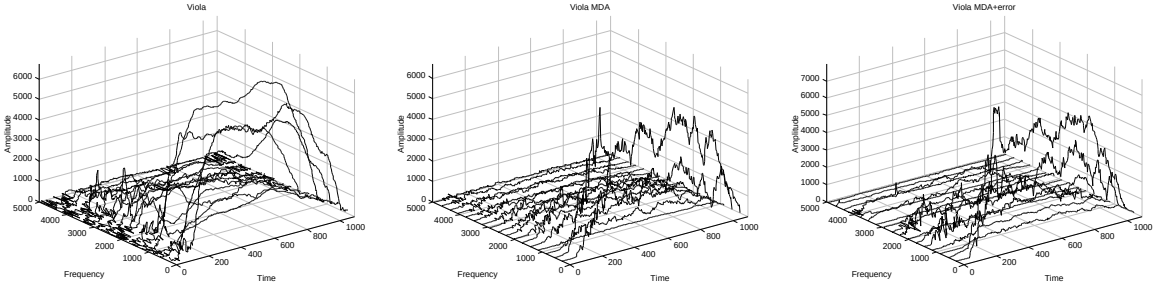


Figure 8.18. Additive parameters for the viola. Original (left) MDA without error (middle) and MDA with error (right).

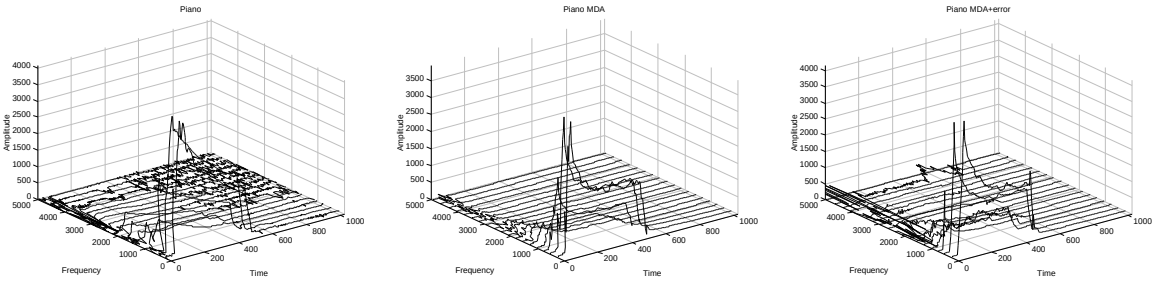


Figure 8.19. Additive parameters for the piano. Original (left) MDA without error (middle) and MDA with error (right).

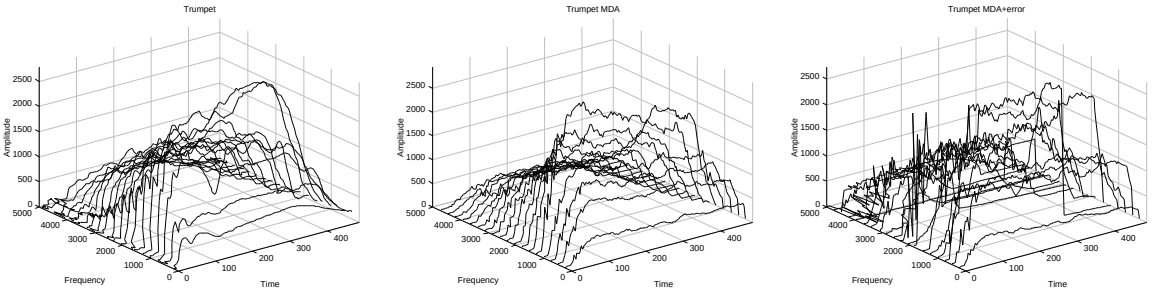


Figure 8.20. Additive parameters for the trumpet. Original (left) MDA without error (middle) and MDA with error (right).

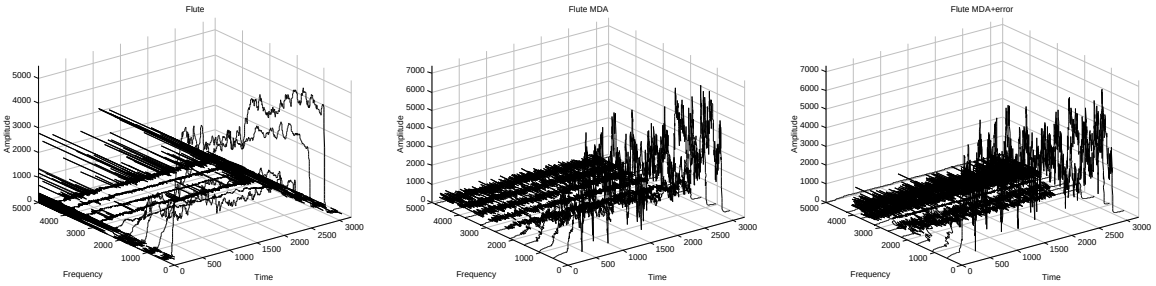


Figure 8.21. Additive parameters for the flute. Original (left) MDA without error (middle) and MDA with error (right).

The trumpet sounds close to the original, the viola and the piano sound different, and the flute is very distorted; indeed, the noise of the flute has a different quality. Still, the sounds are very much recognizable, and the problem can be traced to the fact that the lower stronger partials have a too big noise value. This can perhaps be solved with another noise model, as proposed in 8.4.3, or with different weighting of the noise values. The sound synthesis from MDA parameters with error term is still identifiable, but this model sometimes introduced artifacts, such as high frequency jitter or loose partials, which stick out from the otherwise homogenous sound. Aside from the artifacts, the sound is definitely different without being another instrument, so this method seems promising. Still, the parameters created from the MDA with error need to be limited in some way, so the annoying artifacts do not occur.

8.9. Conclusion

A number of essential timbre attributes are fond in the MDA model. These parameters can be used to resynthesize a sound, as shown in this chapter, but they can also be used to visualize important timbre attributes, as shown in Chapter 9. They can be used as a template in the modification of sounds, as shown in Chapter 10 and they can be used for the classification of musical instrument sounds, as shown in Chapter 11. The sound quality of the resynthesis from the MDA model is evaluated in the listening tests performed in Chapter 12.

Although these parameters are visually close to the HLA parameter set, and the sounds created from the MDA parameters are perceptually close to the HLA sounds, the sound quality is not as good. The problem seems to be the noise parameters, which, when modeled with an exponential curve, often give the wrong value to the lower partials. A polynomial model seems to correspond better to the important noise standard deviation values.

Analysis of the error terms shows that it is often a result of bad parameter estimation, rather than the result of a bad model. Nevertheless, the parameters give a good restitution of the sound, if the analysis is performed correctly. The trumpet, for instance, has a very close resemblance to the original. Further improvements to the models would be a better noise standard deviation model, or a different weight on the different parameters. The MDA model is also improved by using only the lower strong partials in the estimation of the parameters of the model.

The frequencies are well modeled by the fundamental and the inharmonicity, the amplitudes are well modeled by the spectral envelope model as long as there are no formants. The envelope times, percents and curve forms are generally well modeled with the exponential curve, if the weaker, noisy high partials are not used in the estimation of the parameters. Improving the noise model would probably give the best improvement in sound quality.

The MDA model finds a few important parameters for the sound, and while some problems with the parameter estimation exists, the MDA model parameters are believed to be a good guess of the minimum number of parameters necessary to describe a musical instrument sound.

Chapter Nine

9. Instrument Definition Attributes

In this chapter the timbre attributes for many executions of the same instrument are collected in the Instrument Definition Attributes (IDA) model, in half octave bands. This shows clearly the evolutions of the timbre attributes as a function of fundamental frequency. Furthermore, the IDA model can visualize changes from different playing styles, or different intensities. The IDA parameters are assumed to give a complete description of a musical instrument, ranging from the definition of the timbre of one sound, to the evolution of the timbre as a function of note or expression.

The evolution of the timbre attributes is analyzed here as a function of fundamental frequency, intensity, tempo and style. Some simple rules of timbre changes have been found which are helpful when changing, for instance, the pitch of a sound. The IDA model contain all information about a musical instrument, and it can be used to create sounds in the full playing range of the instrument, although the resynthesis quality is not yet satisfactory.

9.1. Introduction

It is not enough to model one sound of a musical instrument to recreate the music of that instrument. It is also important to model the evolution of timbre from the low notes to the high notes of the instrument. Furthermore the evolutions from low velocity to high velocity, from different playing styles, such as *legato* and *staccato* etc., are also important. For this reason, the instrument definition attributes (IDA) model has been introduced. This model keeps the mean of every MDA parameter for each half octave, for each playing style, intensity, etc. The MDA model is presented in Chapter 8. The sound can then be recreated by choosing the note and interpolating the intensity, style, etc. Although further work is needed to achieve an acceptable quality of the recreated sounds, the IDA is useful when the evolution of the timbre attributes are analyzed. If a resynthesis of good quality is needed, the IDA parameters could be derived from the HLA model presented in Chapter 6 instead of the MDA model, but this creates other problems, such as the variable number of partials in the HLA model.

Some indications of the evolution of the timbre attributes can be found in general books on musical acoustics [Backus 1970], [Benade 1990], [Rossing 1990]. Gregory Sandell has a web site [Sandell 1998] with plots of the brightness, irregularity and loudness for different musical instruments, which correlate well with the values found here. The physics of musical instruments [Fletcher *et al.* 1991] may also be of help in evaluating the evolution of the different timbre attributes as a function of fundamental frequency, intensity or other parameters.

This chapter starts with a presentation of the IDA frequency scale in section 9.2. The IDA values are calculated in section 9.3. The different IDA classes are enumerated in section 9.4, and the evolutions of different MDA parameters as a function of fundamental frequency for different instruments are analyzed in section 9.5. The intensity evolution of piano sounds is analyzed in section 9.6, and an analysis of the parameters with two different tempi of the clarinet is presented in section 9.7. The analysis of three different styles of the cello is presented in section 9.8. The sound quality of the resynthesis from the IDA model is discussed in section 9.9. Finally a conclusion is offered.

9.2. Half Octave Bands

The IDA attributes are the same as the MDA attributes, but they are collected for many sounds of one instrument. A number of sounds in the full playing range of a musical instrument are analyzed, and the MDA parameters are created for each sound. The IDA value for each parameter and IDA frequency index is then the mean of the values from the MDAs with fundamental frequency within the corresponding band.

The IDA frequency range is divided into 15 bands in the log2 domain of the fundamental frequency. The bands range from 32 Hz to 4 kHz in half octave steps. All MDA parameters are then searched for each band, and the ones whose fundamental frequency is between the band ± 1/4 octave are used. Each parameter (spectral envelope, frequencies, envelope, noise, etc.), including the partial evolution, of the band is set to the mean of the corresponding parameter of the MDAs used. If no MDAs are found, it is probably outside the playing range of this instrument, and the closest MDA is used.

The MDA parameter values with the fundamental frequency f_0 are added to the IDA step if,

$$\log_2(f_0) \geq (i+9)/2 - \frac{1}{4} \;\&\; \log_2(f_0) < (i+9)/2 + \frac{1}{4} \qquad \textbf{(9.1)}$$

This means that the frequency range for each IDA index step is

1 (26.9 Hz to 38.1 Hz)	9 (430.5 Hz to 608.9 Hz)
2 (38.1 Hz to 53.8 Hz)	10 (608.9 Hz to 861.1 Hz)
3 (53.8 Hz to 76.1 Hz)	11 (861.1 Hz to 1217.7 Hz)
4 (76.1 Hz to 107.6 Hz)	12 (1217.7 Hz to 1722.2 Hz)
5 (107.6 Hz to 152.2 Hz)	13 (1722.2 Hz to 2435.5 Hz)
6 (152.2 Hz to 215.3 Hz)	14 (2435.5 Hz to 3444.3 Hz)
7 (215.3 Hz to 304.4 Hz)	15 (3444.3 Hz to 4871.0 Hz)
8 (304.4 Hz to 430.5 Hz)	

The IDA frequency scale can be seen in figure 9.1. The scale range is divided into half-octave steps.

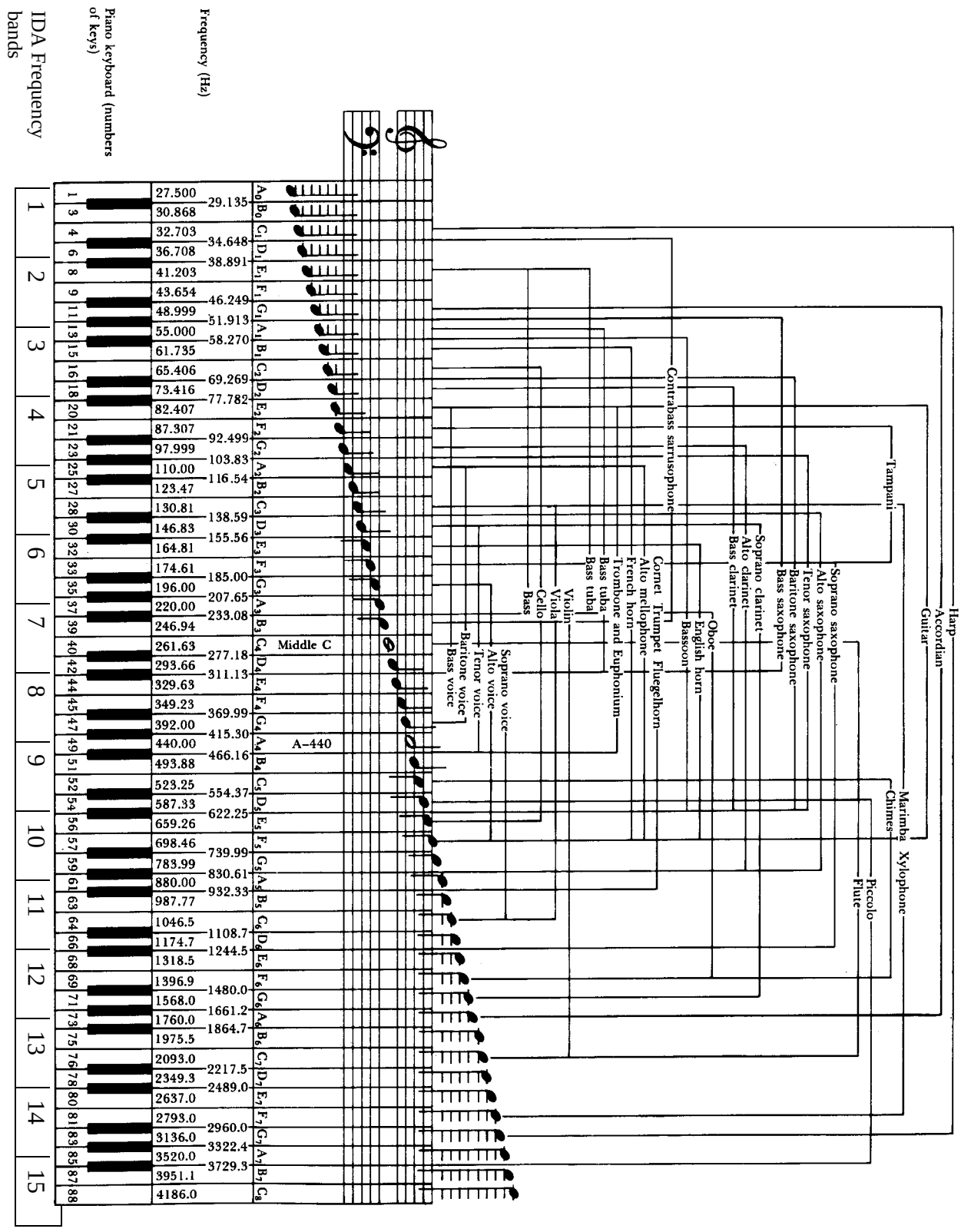


Figure 9.1. Approximative IDA frequency bands for different musical instruments (picture taken from [Lindsay *et al.* 1977]).

The IDA parameters become more stable, the more sounds there are in each frequency band. Initially it was believed that the IDA could remedy some analysis errors, i.e. the MDA parameter errors might be great, but the mean of the errors would be zero. This seems true for the visualization of the timbre attributes, but no such conclusion can be made for the resynthesis, maybe because not enough sounds have been used in the creation of the IDAs, or possibly because the errors in the MDA model are very correlated.

9.3. IDA Parameter Calculation

The IDA model parameters are derived from the MDA model parameters, which in turn are extracted from the HLA model parameters. The HLA model parameters are calculated from the additive parameters, which are analyzed from the sampled sound.

The additive parameters are calculated using the LTF analysis method as explained in Chapter 4. The initial frequencies used in the analysis are the frequencies found in Chapter 3, with one important exception. The fundamental frequency estimation is given the note of the sound to analyze. The initial frequency search is therefore simplified: first find the frequency differences that are close to the given note, then do the stretched frequencies curve fit, and finally look for spurious frequencies. It was necessary to use this method in order to eliminate the influence of the fundamental frequency estimation error.

The HLA parameters are calculated from the additive parameters as explained in Chapter 6 and the MDA parameters are calculated from the HLA parameters as explained in Chapter 8.

The value for each IDA frequency band is set for each parameter to the mean of the corresponding values of all MDA which have the fundamental frequency in the frequency band.

$$IDA_k = mean(MDA(f_0 \quad band_k)) \qquad \textbf{(9.2)}$$

If there is no MDA with fundamental frequency in a frequency band, the closest MDA is used. This ensures that all IDA values always are set. If there is only one available MDA for the creation of an IDA class, then the IDA parameters are equivalent to the MDA parameters, although lacking the fundamental frequency.

The parameters of the plots in this chapter are the values of the fundamental, recreated from the fundamental values (v_0 in equation (8.7) on page 109) and the partial evolution value.

9.4. IDA Classes

A separate IDA is created for each playing style, intensity or other class, since not all instruments have the same number of playing styles, and it would be difficult to organize them in the same manner for all instruments.

Typical classes are the different intensities of an instrument, such as *piano*, *mezzo forte*, or *forte*. Other classes are the different playing styles of an instrument, such as *legato*, *staccato*, etc., and the tempo of the execution. Furthermore, it is interesting to classify the instruments in the physical dimensions of the gestures of the instrumentalist, such as the speed or the position of the bow in the violin.

Often, fewer samples are available for some of the IDA classes than for other. Then the attributes from the largest class should be used in case no MDA is available for the target class, with a difference value added. For instance, if both the target class[a] and the largest class[b] have MDAs at frequency band *k*, but the IDA values are wanted from class[a] and frequency band *j*, where only class[b] has MDA values. Then the resulting values could be the values of class[a] plus the difference between the values from class[b] and class[a], in the closest frequency band *k* where both classes have MDA parameters,

$$IDA_j^a = HLA^b(f_0 \quad band_j) + HLA^a(f_0 \quad band_k) - HLA^b(f_0 \quad band_k) \qquad \textbf{(9.3)}$$

If no common frequency band exists, other more elaborate schemes could be found, but this is beyond the scope of this work. Generally, it probably makes more sense to use the values from the closest frequency band where values from MDAs exist.

9.5. Fundamental Frequency Evolution

In this section, the evolutions of all timbre attributes are analyzed as a function of fundamental frequency. The analysis is done on five instruments, the piano, the violin, the clarinet, the flute and the soprano. All instruments have sounds from the normal playing range of the instrument. These are the same sounds used in the classification in Chapter 11 and in the listening tests in Chapter 12.

9.5.1. Spectral Envelope Evolution

The spectral envelope parameters are brightness, tristimulus, the odd/even relation, irregularity and amplitude. The spectral envelope model parameters are plotted in figure 9.2 for the piano, in figure 9.3 for the violin, in figure 9.4 for the clarinet, in figure 9.5 for

the flute and in figure 9.6 for the flute. The parameters which are plotted are the brightness in Hz, that is the partial index brightness multiplied by the fundamental frequency (top left), the odd value (bottom left), the tristimulus 1 and 2 (top right) and the irregularity (bottom right). All x-axes, except for the tristimulus are in IDA frequency band index. Two plus signs at the x-axis depict the fundamental frequency range of the instrument of the plot. The tristimulus plots do not have an IDA frequency band axis, but the curve can be followed from the lowest frequency '+' to the highest frequency 'o'.

The brightness' in partial index for 5 instruments are plotted in figure 9.8 and the amplitudes for all five sounds are plotted in figure 9.7.

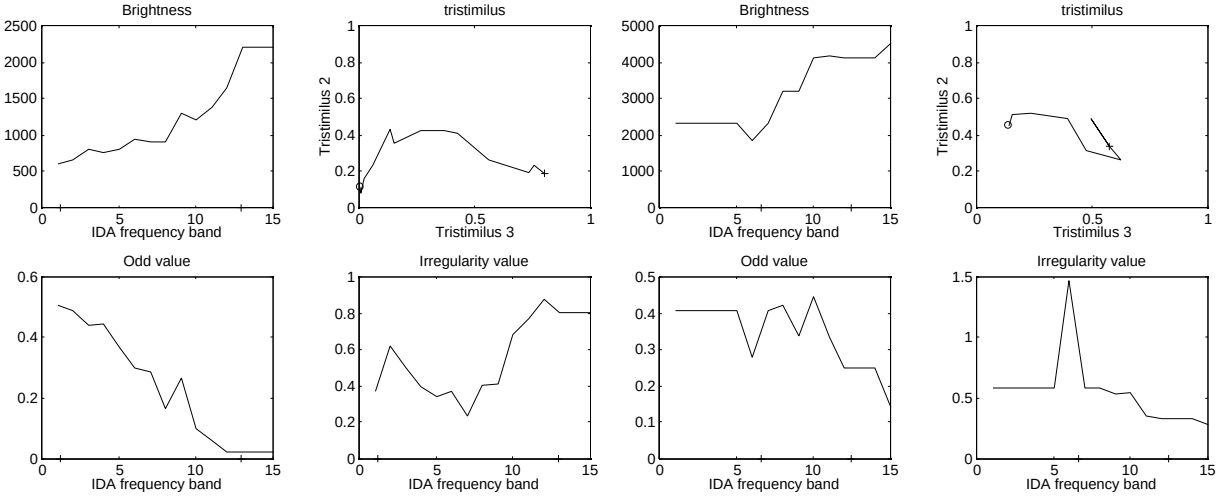


Figure 9.2. Spectral parameters for the piano.

Figure 9.3. Spectral parameters for the violin

The spectral envelope attributes changes seem very important when changing the fundamental frequency. An initial analysis of the spectral envelope attribute changes reveals some properties of these attributes, which corresponds to most of the instruments. The attributes that seem to have an simple law associated with the fundamental frequency change are brightness, tristimulus, odd and amplitude.

The partial index brightness multiplied by the square octave index is roughly constant for most instruments. This is a handy guideline, if the pitch of a sound is modified, and gross spectral envelope effects avoided. The amplitude divided by the log of the octave is also more or less constant over the full fundamental frequency range for most instruments. Tristimulus 1 divided by the octave, and tristimulus 2 multiplied by the octave are fairly constant for all sounds, as is the odd times the square octave. All these observations are valid for most of the instruments to some degree.

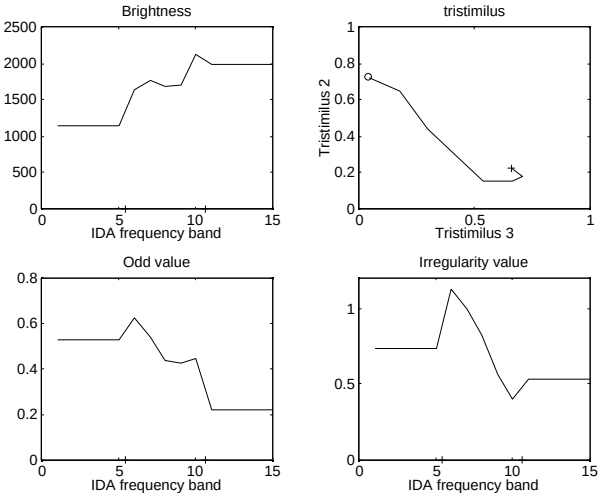


Figure 9.4. Spectral parameters for the clarinet.

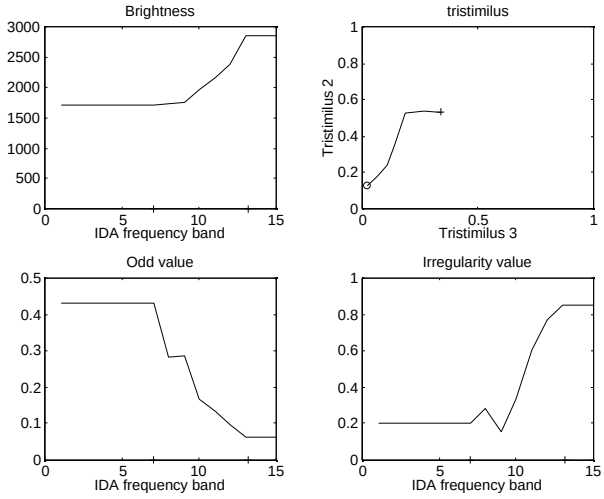


Figure 9.5. Spectral parameters for the flute.

Except perhaps for brightness and amplitude, these rules are not valid in all cases. They can be used, though, if the pitch of a sound is to be changed, and no other information of the timbre attribute changes for that pitch is available.

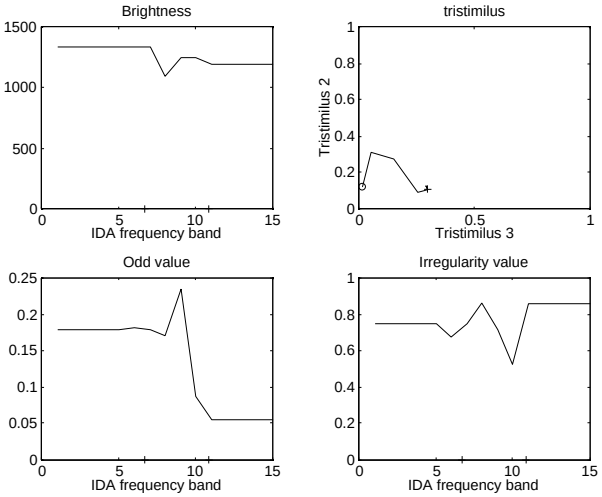


Figure 9.6. Spectral parameters for the soprano. Brightness (top left), tristimulus (top right), odd (bottom left) and irregularity (bottom right).

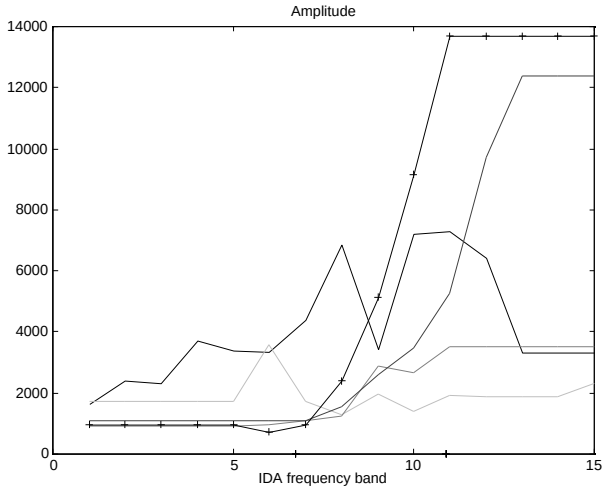


Figure 9.7. Amplitude for the 5 instruments. Piano (solid), violin, (dotted), clarinet, (dashdotted), flute, (dashed) and soprano (+-solid).

All instruments except the soprano have frequency brightness rising with the fundamental frequency. The piano has a rising brightness curve, going from almost 500 Hz at the lowest notes, to above 2000 Hz at the highest notes. This is in strong contrast to the soprano voice, which has a stable brightness at about 1200 Hz. This can be explained by the fact that the soprano needs to keep the same formantic structure all the time, since the same vowel is used for all notes.

Nevertheless, the frequency brightness rise is not so dramatic as the tristimulus 1 rise for the piano, so the fundamental has more relative strength for the high notes.

The partial index brightness shown in figure 9.8 makes it clear that most instruments have most of the amplitude in the fundamental for the highest frequencies, since brightness always tends towards one for the highest pitches.

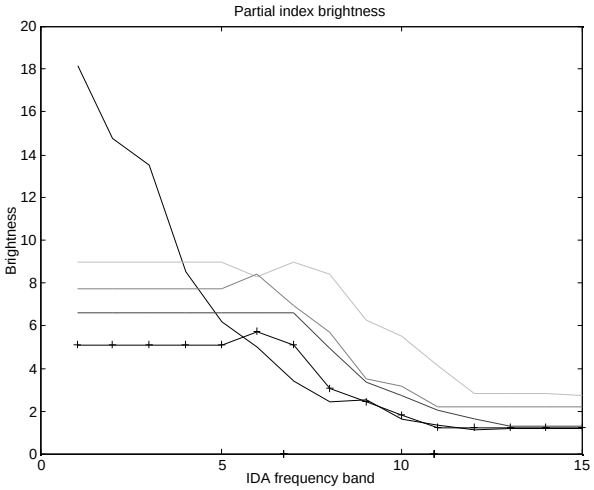


Figure 9.8. Partial index brightness for 5 instruments. Piano (solid), violin, (dotted), clarinet, (dashdotted), flute, (dashed) and soprano (+-solid).

The tristimulus general trend is towards the fundamental corner. This is especially true for the flute. The violin has a fairly constant tristimulus 2 value, and the clarinet has a rising tristimulus 2 value. Common to all sounds is a falling tristimulus 3 value, indicating weaker high-frequency amplitudes.

The tristimulus 1 value for the piano goes from nearly zero for the low notes to close to one for the high notes. It is also interesting to observe the jump in tristimulus 1 at about 2/3 of the scale. This might be explained by the difference in string quality, the number of strings or the string coating.

It is also interesting to observe the tristimulus curves for the soprano: first tristimulus 2 rises, and then tristimulus 1. This might be explained by the place of the first formant: for the low notes the first formant is placed above the 4th partial, but as the fundamental frequency rises, the 4th, 3rd and finally 2nd partial are amplified by the first formant. Finally the fundamental is placed in the formant region, which is consequently forced to rise with the fundamental. See [Sundberg 1987] for a further explanation of this phenomena.

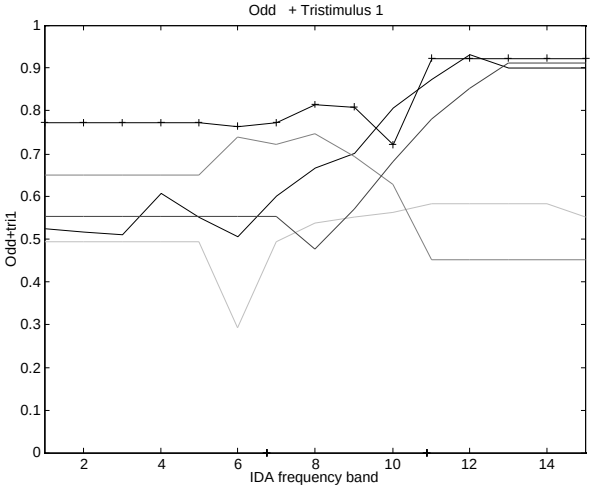


Figure 9.9. Odd plus tristimulus 1 for the five instruments. Piano (solid), violin, (dotted), clarinet, (dashdotted), flute, (dashed) and soprano (+-solid).

The odd value is falling with the fundamental frequency for all five instruments. This seems to be more because the fundamental amplitude is rising than because the odd/even relation is changing. This can be verified by adding the tristimulus 1 to the odd.

The odd plus tristimulus 1 for the five instruments are plotted in figure 9.9. The soprano has the highest values, followed by the clarinet. The violin has the lowest values. The piano and flute values rise with the fundamental frequency. These values are more stable than the odd values.

The irregularity value is rising with the fundamental frequency for the piano, flute, and to a lesser degree, the soprano. This is correlated with the tristimulus 1 value. When only the fundamental has strong amplitude, the irregularity is by definition 1. The clarinet and the violin have, by contrast, a falling irregularity, which to some degree is caused by the weak fundamental of the low notes, but also by the general irregular shape of the spectral envelope of the low notes for these two instruments, due to the weak modes.

Amplitude is strongly correlated with the fundament frequency for all instruments, except the violin, which has a more stable amplitude. The piano seems to have two regions for the amplitude, which is probably explained by the shift of string quality, or string number of each note. The clarinet also seems to have two regions for the amplitude; this is probably explained by the shift of mode for the high frequencies [Fletcher *et al.* 1991].

9.5.2. Frequency Parameter Evolution

The MDA model frequency attributes consist of the fundamental frequency, and the harmonicity. The IDA model does not save the fundamental frequency of each MDA, since it is inherent in the model: the IDA bands indicate the log of the fundamental frequency. Therefore, the only frequency parameter in the IDA model is the inharmonicity.

The inharmonicities of the five sounds are shown in figure 9.10. The inharmonicity is rather noisy, especially for the high frequencies, because of the small number of partials, but it definitely seems to rise with the fundamental frequency for the piano. Remember that the center frequency of band 12 is 1500 Hz, which means that there is only a maximum of 10 partials with the sampling rate used here (32 kHz).

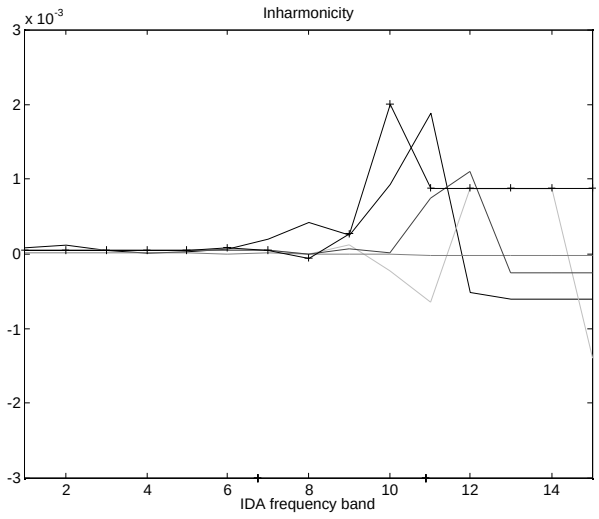


Figure 9.10. Inharmonicity for the 5 instruments. Piano (solid), violin, (dotted), clarinet, (dashdotted), flute, (dashed) and soprano (+-solid).

If the frequencies of these partials are estimated poorly, the resulting inharmonicity value could be important, even though there was no inharmonicity.

9.5.3. Envelope Evolution

The envelope attributes are the envelope times, curve forms, and percents. The most interesting envelope times are the attack time and the release time. The attack and release envelope parameters are shown in figure 9.11 for the piano, in figure 9.12 for the violin, in figure 9.13 for the clarinet, in figure 9.14 for the flute and finally in figure 9.15 for the soprano. The left three plots are the attack time (top), the end of attack percent, and the attack curve form (bottom). The three right plots are the corresponding release parameters. The sustain curve form and length for the five instruments are shown in figure 9.16 and the start curve form and percents are shown in figure 9.17.

The attack time can be divided up into slow attacks, flute, clarinet and soprano, at about 100 mS, and fast attacks, violin and piano, at below 50 mS. Most instruments seem to have faster attacks for higher notes, but this is very clear for the flute, and especially for the piano, which goes from about 80 mS for low notes to about 20 mS for high notes. The violin seems to have a stable attack time, independent of the fundamental frequency.

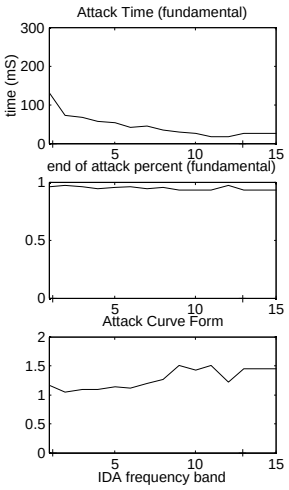
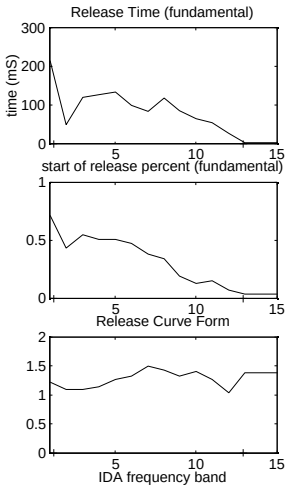
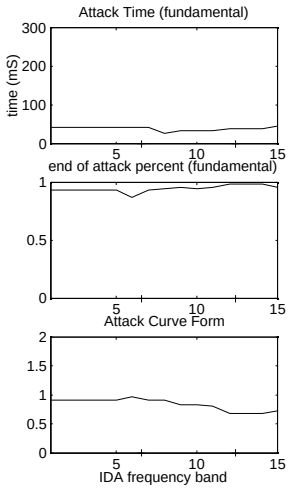


Figure 9.11. Envelope parameters for the piano





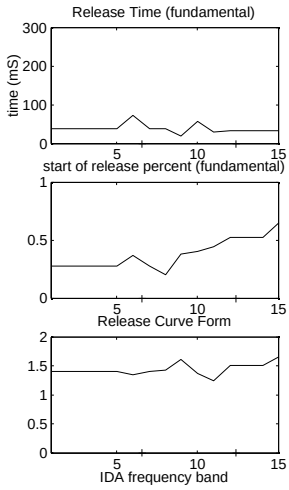


Figure 9.12. Envelope parameters for the violin.

The release time values are slightly noisier, probably due to the decay/sustain model of the analysis. Nevertheless, the general trend seems to be that release is more independent of the fundamental frequency than the attack times. The violin, flute and soprano have relatively stable release times, and only the piano and the clarinet have a decaying release time, which for the clarinet is about the same as the attack time. The piano has slower

release than attack times, whereas the violin, flute and soprano have faster release than attack times.

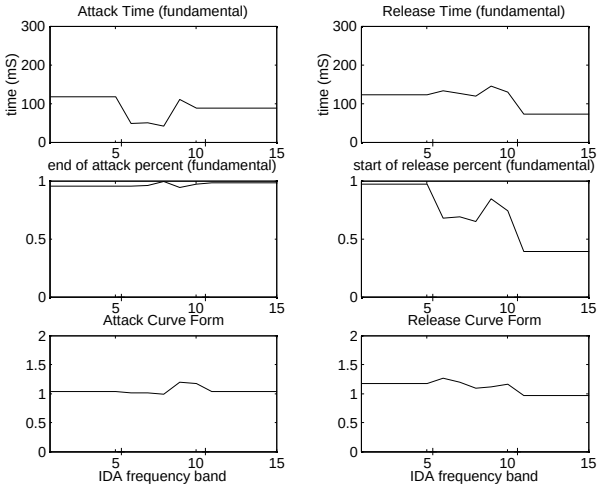


Figure 9.13. Envelope parameters for the clarinet.

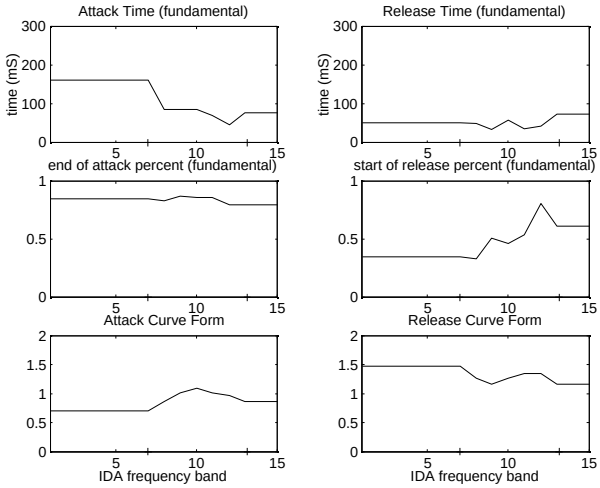


Figure 9.14. Envelope parameters for the flute.

The curve forms seem close to linear for most attacks and releases. The attack curve forms can be divided into logarithmic for the piano, clarinet and soprano, and exponential for the violin and flute. The release curve forms seem exponential for all instruments.

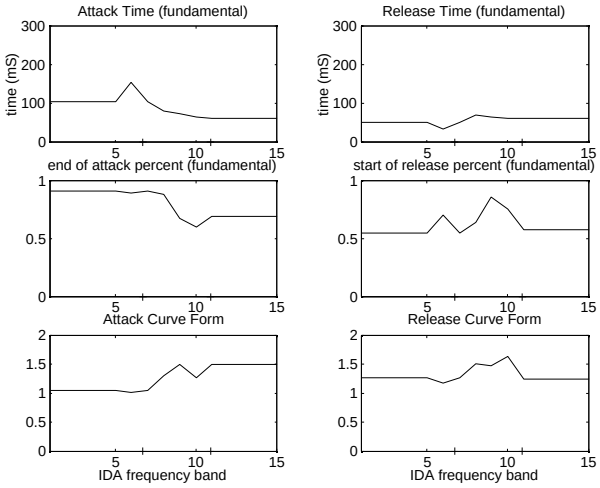


Figure 9.15. Envelope parameters for the soprano. Attack (left) and release (right). Time (top), percents (middle) and curve form (bottom)

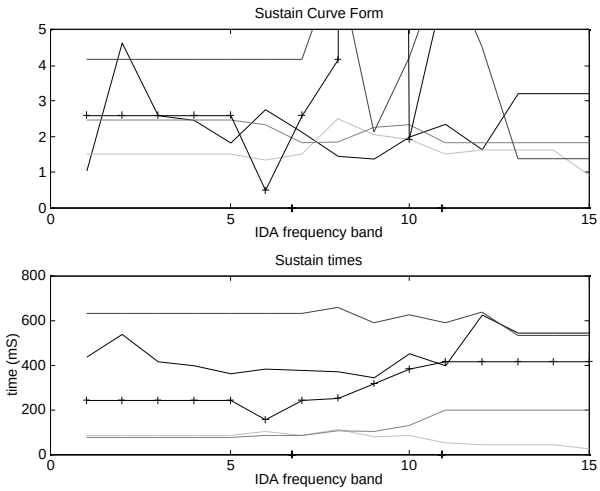


Figure 9.16. Sustain Curve form (top) and sustain length (bottom) for the 5 instruments. Piano (solid), violin, (dotted), clarinet, (dashdotted), flute, (dashed) and soprano (+-solid).

The sustain curve form values, which can be seen in figure 9.16 (top), are always above 1. Since the start of release percents is lower than the end of attack percents in almost all cases, this indicates an exponential decay maybe because all instruments are played in a percussive style.

It is hard to make other observations on the curve form, since it is very dependent on the correct estimation of the envelope times. The attack curve form value seems to rise slightly with the fundamental frequency for some instruments, indicating a more logarithmic attack for higher fundamental frequencies.

The sustain times in figure 9.16 (bottom) are rather constant. The sounds are short, especially the flute and the clarinet. Since all sounds from one instrument come from the same recording session, they are all of the same duration and the invariance of the sustain times shows the success of the sustain time estimation. Only the soprano has sustain time rising with the fundamental frequency.

The clarinet and soprano start segment curve forms, seen in figure 9.17, decrease with the fundamental frequency, almost reaching zero. This gives very exponential curves, which rise abruptly at the start of the attack.

The slope envelope detection method should not normally be sensitive to noise, and an inspection of the amplitudes of the additive partials of the relevant sounds reveals that they are indeed exponential in the start segment, more rounded for the clarinet, and rising very abruptly for the soprano.

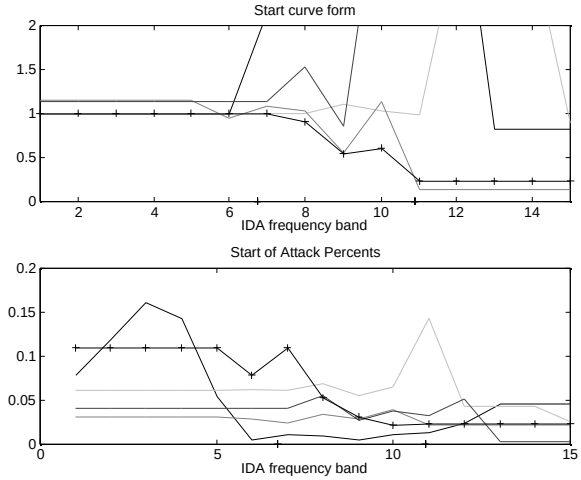


Figure 9.17. Start curve form (top) and start of attack percents (bottom) for the 5 instruments. Piano (solid), violin, (dotted), clarinet, (dashdotted), flute, (dashed) and soprano (+-solid).

9.5.4. Noise Evolution

The noise attributes are the standard deviation, the filter coefficient and the correlation of the irregularity on the amplitude of the partials, the shimmer, and of the irregularity on the frequency of the partials, the jitter.

The noise parameters are plotted in figure 9.18 for the piano, in figure 9.19 for the violin, in figure 9.20 for the clarinet, in figure 9.21 for the flute, and finally in figure 9.22 for the soprano. The left three plots are shimmer parameters and the right three plots jitter parameters. The top plot is the standard deviation, the middle plot is the filter coefficient, and the bottom plot is the correlation.

The correlation is measured between the fundamental and the first partial.

The attack noise standard deviation for shimmer (top) and jitter (bottom) are shown in figure 9.23.

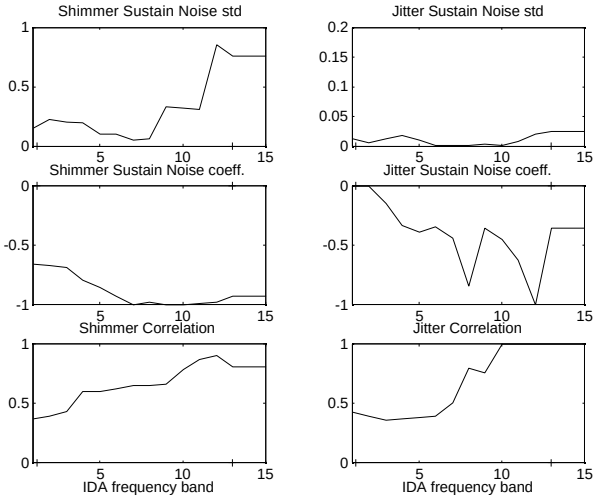


Figure 9.18. Noise parameters for the piano.

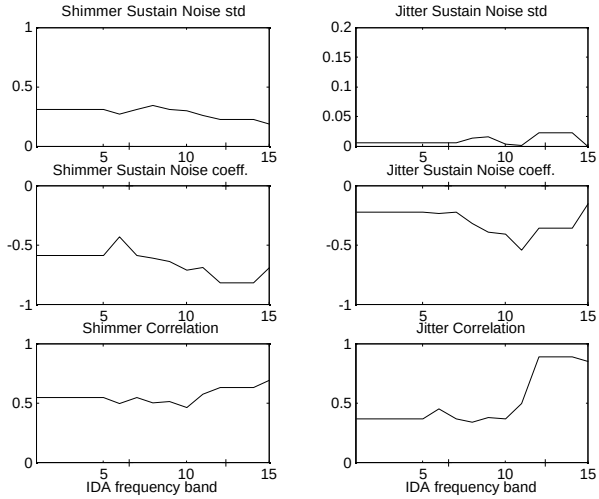


Figure 9.19. Noise parameters for the violin.

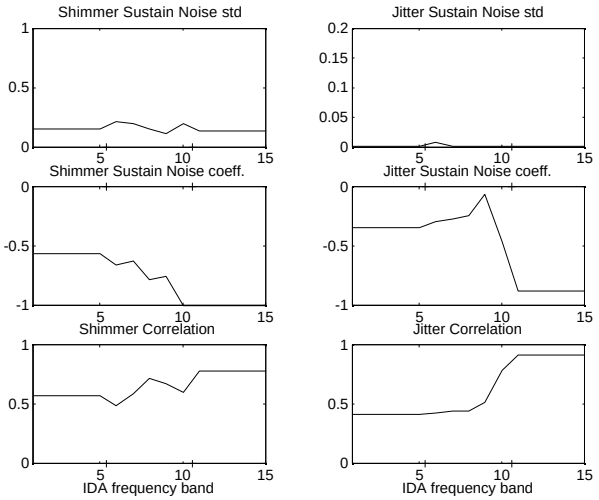


Figure 9.20. Noise parameters for the clarinet.

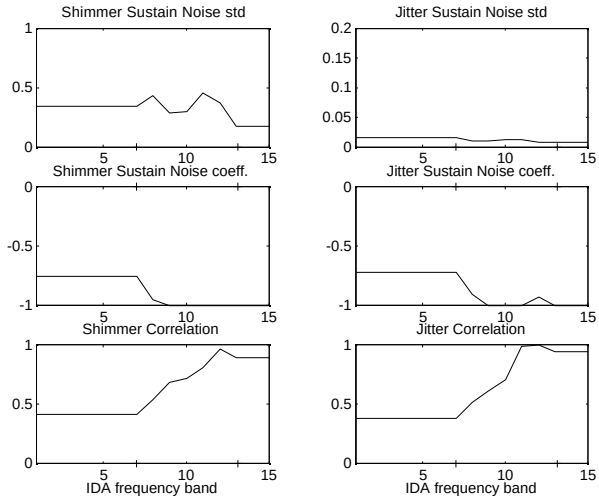


Figure 9.21. Noise parameters for the flute.

The noise attributes seem to be the timbre attributes that are causing most disturbances in the resynthesis of the MDA models. Bad analysis of the envelope values yields bad shimmer values, and bad analysis of the fundamental frequency, or inharmonicity yields bad jitter values. Furthermore, even small glissando or vibrato values can ruin the analysis of the frequencies.

Nevertheless, most noise values are stable and trustworthy. The shimmer and jitter are normalized by the amplitude and frequency, and the standard deviation (std) is generally placed between 0 and 1. The shimmer std is often close to 0.3, whereas the jitter std generally is well below 0.1.

The shimmer standard deviations are rather stable for all instruments except the piano. The flute and violin have higher shimmer than the clarinet and soprano. The clarinet has very low shimmer. The piano has a rising shimmer standard deviation. Since the filter coefficients at the same time tends towards -1, this might be caused by a bad curve form model or fitting, but it could also be attributed to the beating of mistuned strings.

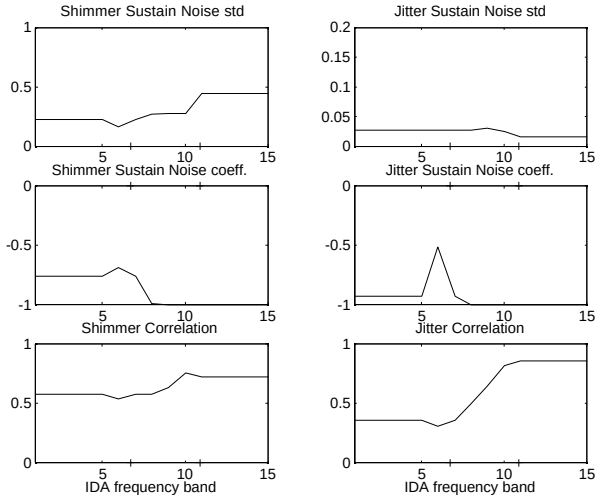


Figure 9.22. Noise parameters for the soprano. Shimmer (left) and jitter (right). Standard deviation (top), filter coefficient (middle) and correlation (bottom).

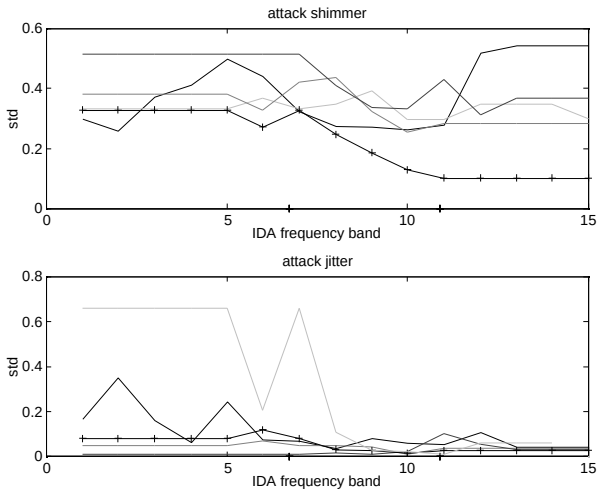


Figure 9.23. Attack shimmer (top) and jitter (bottom)) for the 5 instruments. Piano (solid), violin, (dotted), clarinet, (dashdotted), flute, (dashed) and soprano (+-solid).

The jitter standard deviation is around 0.01 for the piano and the violin. The clarinet has almost no jitter std, whereas the flute and especially the soprano have high jitter std.

The filter coefficients indicate the level of low-pass filter slope of the noise. Generally, the filter coefficient value -1 indicates bad curve fitting; this can be seen also on the shimmer for the high notes of all instruments except the violin. The high notes of the soprano and the flute have filter coefficient value -1 for both the shimmer and the jitter. These jitter values are probably caused by the relatively low frequency vibrato on many of the notes.

The jitter filter coefficients seem to be higher than the shimmer filter coefficients in most cases. This translates into more energy in the high frequencies of the jitter. The combination of high standard deviation and low filter coefficient values for the jitter generally yields bad sound quality, with slow random variations on the frequencies.

The attack noise standard deviations for shimmer (top) and jitter (bottom) are shown in figure 9.23. The attack shimmer std generally falls with the fundamental frequency. The soprano has the lowest attack shimmer std, followed by the piano and the three other instruments close together. The violin has very large jitter std for the low frequencies.

The correlation between the fundamental and the first partial generally rises with the fundamental frequency, reaching almost one in many cases. The shimmer and jitter correlation values are very similar for all instruments. This is surprising, since the envelope curve form model normally should give higher shimmer correlation.

9.6. Loudness

Loudness is another important timbre attributes. Loudness is often noted in music notation with terms like *mezzo forte, piano, forte,* etc. In this paragraph, the different loudnesses of a Yamaha Disklavier with different MIDI [IMA 1983] velocities are analyzed. The disklavier is an acoustic grand piano with an added MIDI control unit, which permits the recording of MIDI data, and the control through MIDI data. Three different MIDI velocities have been recorded in the full playing range, 40 (*piano*), 72 (*mezzo forte*) and 104 (*forte*). The *mezzo forte* sounds are the same as the piano sounds in section 9.5.

The complete playing ranges of three different velocities of the piano are shown in the following figures. Four plots are combined in all figures. The solid lines show the IDA values for all the sounds, the dotted lines the values from the *piano* (MIDI velocity 40) sounds, the dashdotted lines the values for the *mezzo forte* (MIDI velocity 72) sounds and the dashed lines the IDA values for the *forte* (MIDI velocity 104) sounds. Figure 9.24 shows the spectral envelope, figure 9.25 the maximum amplitude, figure 9.26 the inharmonicity, figure 9.27 the envelope parameters, figure 9.28 the sustain curve form values, and figure 9.29 the noise parameters for the different loudnesses of the piano.

The spectral envelope is the attribute that a priori changes the most with loudness. This holds true here, but the first observation is the high correlation of the different loudnesses for all attributes.

9.6.1. Spectral Envelope Parameters

The spectral envelope values for the different loudnesses of the piano are shown in figure 9.24. Brightness is top left, tristimulus is top right, odd is bottom left and irregularity is bottom right. The amplitudes are shown in figure 9.25 for all loudnesses.

Brightness is larger for the *forte* sounds than for the *piano* sounds. Furthermore, the total value is very close to the *mezzo forte*, indicating that the *mezzo forte* sounds are indeed in the middle between the *piano* and the *forte* sounds. The *piano* sounds have less

odd value, which correlates with the fact that they have less brightness and thus more energy in the fundamental. The *piano* sounds also have more tristimulus 2, probably for the same reason.

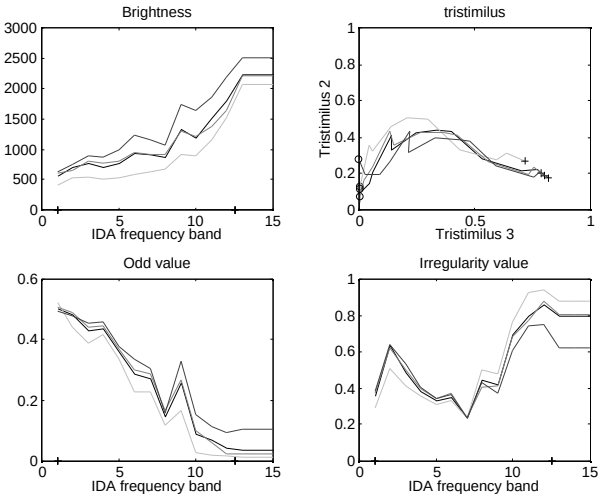


Figure 9.24. Spectral Envelope parameters for three different loudnesses for the piano. All loudnesses (solid), *piano* (dotted), *mezzo forte* (dashdotted) and *forte* (dashed).

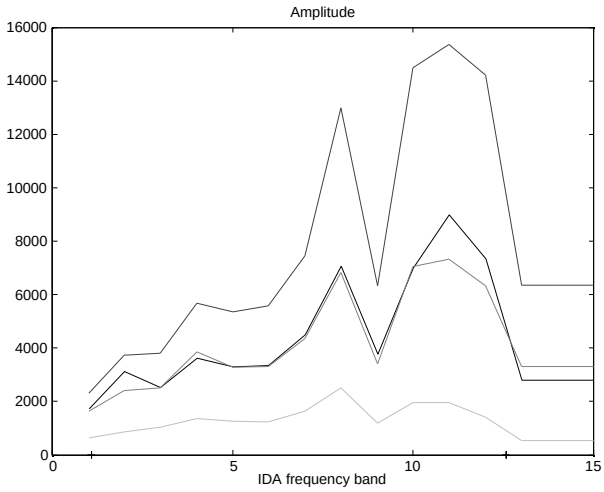


Figure 9.25. IDA fundamental amplitude for three different loudnesses for the piano. All loudnesses (solid), *piano* (dotted), *mezzo forte* (dashdotted) and *forte* (dashed).

The irregularity shape is exactly the same for all loudnesses. The reason for this shape is not clear at this point, but it could be related to the change of quality and number of the strings. The *piano* sounds seem to have more irregularity at the high frequencies, but this could perhaps be attributed to the difficulty of analyzing weak signals.

The amplitudes shown in figure 9.25 have the same shape for all four curves. The *forte* is stronger than the *piano*, of course, and the total curve is very close to the *mezzo forte*, which again indicates that the *mezzo forte* is exactly between the *forte* and the *piano* loudnesses.

9.6.2. Frequency Parameters

Inharmonicity for the different loudnesses of the piano can be seen in figure 9.26.

The inharmonicities also have the same shape for all curves. The *piano* sounds have a more irregular shape, which again could be attributed to the difficulty of analyzing weak signals. The inharmonicity increases with the fundamental frequency for the *mezzo forte* and the *forte* sounds.

The estimated values show clearly that the *piano* sounds have less inharmonicity than the *forte* sounds in the high notes.

Whether this is an effect of the analysis, which often is less reliable for the weak sounds, is unclear at this stage. The high fundamental frequency inharmonicity is also less reliable than the low fundamental frequency inharmonicity, due to the fewer partials to calculate the inharmonicity from.

The low fundamental frequency inharmonicities have no reliable difference among the loudnesses

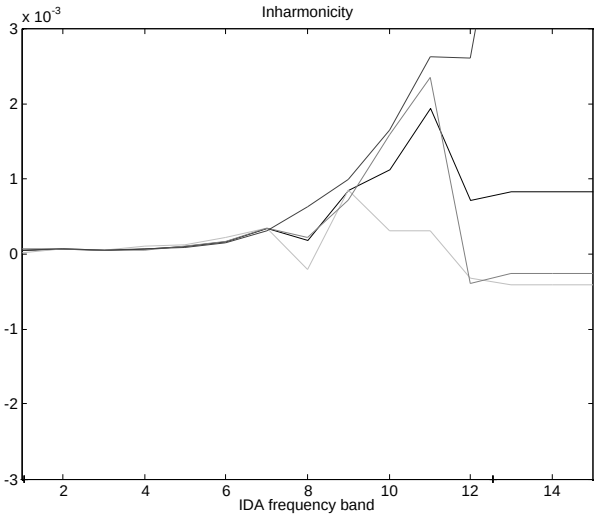


Figure 9.26. Inharmonicity for three different loudnesses for the piano. All loudnesses (solid), *piano* (dotted), *mezzo forte* (dashdotted) and *forte* (dashed).

9.6.3. Envelope Parameters

The envelope parameters for the different loudnesses of the piano can be seen in figure 9.27. The left three plots are the attack time (top), the attack percents (middle) and the attack curve forms. The right three plots give the corresponding values for the release.

The curve form (top) and the length (bottom) of the sustain for the different loudnesses of the piano are shown in figure 9.28. The sustain curve form values are reliable, since the start of release percents are much lower than the end of attack percents.

The attack times are remarkably similar for all loudnesses. They decrease linearly from around 100 mS to below 20 mS for all loudnesses. The release times first increase in the very low frequencies for all loudnesses, after which they decrease linearly, from almost 200 mS to around 20 mS, in the normal playing range. The highest notes have more noise, but the *mezzo forte* and *forte* values, which are less sensitive to noise, continue to decrease.

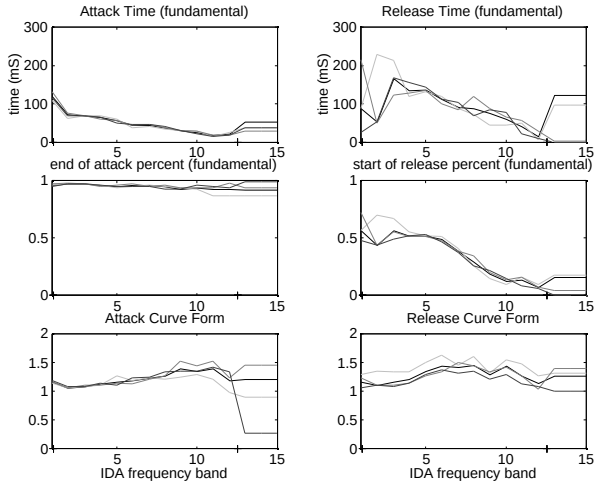


Figure 9.27. Envelope parameters for three different loudnesses for the piano. All loudnesses (solid), *piano* (dotted), *mezzo forte* (dashdotted) and *forte* (dashed).

The attack percent values are rather stable close to 1, but the release percents decrease with the fundamental frequency, which indicates a faster decrease rate for the highest notes. The attack curve form is close to linear, getting slightly more logarithmic at higher frequencies and with more noise, or irregularity, at the high frequencies. The *forte* sounds are more logarithmic than the *piano* sounds at high frequencies.

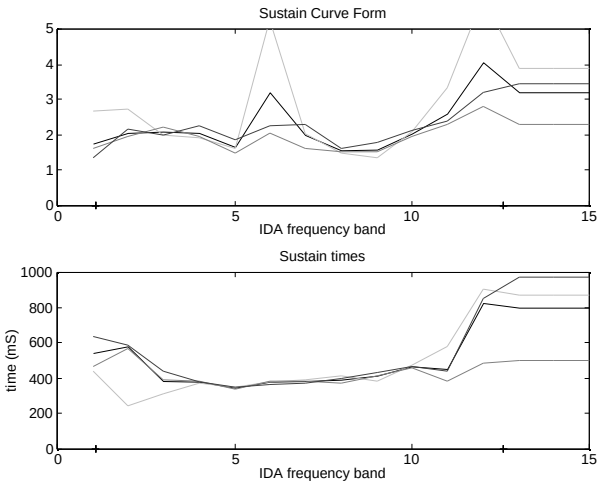


Figure 9.28. Sustain curve form values (top) and sustain length (bottom) for three different loudnesses for the piano. All loudnesses (solid), *piano* (dotted), *mezzo forte* (dashdotted) and *forte* (dashed).

The release curve form values seem dependent on the loudness for the full fundamental frequency range.

The *piano* sounds have a more exponential release than the *forte* sounds. The release curve form also rises with the fundamental frequency up to about midrange after which it falls again, indicating a more exponential release for the midrange sounds. The curve form values are rather constant, but rising for the high notes. The values are always above 1, which indicates an exponential decay, more exponential for high frequencies. The noises on the sustain curves, especially the peak in the midrange, are in part due to noise from the *piano* sounds.

The sustain lengths are rather constant and, since both the attack and release times decrease with the fundamental frequency, indicate shorter high fundamental frequency sounds.

9.6.4. Noise Parameters

The sustain noise parameters for the different loudnesses of the piano are shown in figure 9.29. The shimmer parameters are shown in the left column and the jitter parameters in the right column. The top plots show the standard deviation, the middle plots show filter coefficients, and the bottom plots show correlation. The combinations of all loudnesses are shown as a solid line, the *piano* values are dotted, the *mezzo forte* values dashdotted and the *forte* values dashed.

The noise parameter curves are also very correlated with the different loudnesses. The shimmer standard deviation value starts at around 0.4, then it decreases to just above 0.1 at

around 200 Hz, after which it increases to almost 1 at 1.5 kHz. This behavior is the same for all loudnesses, although the *piano* sounds seem to have a slightly lower shimmer std for low frequencies.

The jitter standard deviation values also have the same shape for all loudnesses, but here the *piano* sounds have more jitter std than the *forte* sounds in the high notes.

The shimmer filter coefficient decreases from around -0.8 to almost -1. The jitter filter coefficients also decrease, with some interruptions which are believed to be noise, from around -0.1 to around -0.6.

Shimmer and jitter correlations increase with the fundamental frequency for the *mezzo-forte* sounds, but decrease with the fundamental frequency for the *piano* and *forte* sounds.

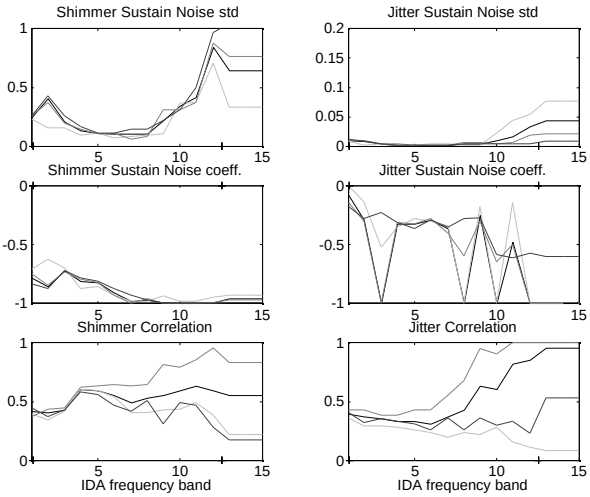


Figure 9.29. Noise parameters for three different loudnesses for the piano. All loudnesses (solid), *piano* (dotted), *mezzo forte* (dashdotted) and *forte* (dashed).

Attack shimmer and jitter standard deviations are shown in figure 9.30. The shimmer is rather constant, with a std between 0.3 and 0.5. The jitter exhibits large variations. Since the large jitter std values occur for the *mezzo forte* and *forte* sounds, it seems that *forte* sounds have more transient behavior than *piano* sounds. However, this does not seem to be true for the low mid range. The *piano* sounds always have low jitter std, despite the fact that weak sounds are generally harder to analyze.

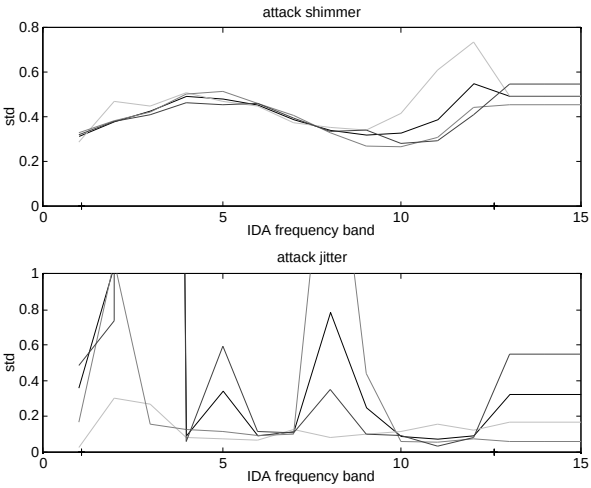


Figure 9.30. Attack shimmer (top) and jitter (bottom) std. All loudnesses (solid), *piano* (dotted), *mezzo forte* (dashdotted) and *forte* (dashed).

9.6.5. Loudnesses conclusions

Three conclusions can be drawn from the analysis of the IDAs from the piano sounds with different loudnesses. First of all, the difference in loudness translates mostly into a difference in the spectral envelope parameters, and especially into a difference in the brightness and amplitudes. Secondly, most curves are very similar across loudnesses, even those suspected to be rather noisy. Third, the *piano* IDA parameters are generally more noisy than the *forte* parameters. The *piano* sounds also have more noise than the *forte* sounds, and the *forte* noises are more correlated than the *piano* sounds. Additional changes in the IDA model parameters for a change in the loudness of a piano include a change in release curve form, *piano* sounds having a more exponential release. *Piano* sounds seem to have less inharmonicity and *forte* sounds have more jitter in the attack.

In conclusion, the IDA values seem eminently suitable for the analysis of different loudnesses. The IDA values are generally very stable, and the differences among loudnesses are very clear.

9.7. Tempo

Tempo is another important expression parameter. Tempo is generally written in scores with terms such as *moderato*, *allegro*, etc. Here, two performances with different tempi of the clarinet are analyzed. The tempi are *allegro* and *moderato*. The loudnesses are a mix of *piano* and *forte* and the playing style is *staccato*. The full playing range of the clarinet is available for the two tempi. In the following figures the combined IDA values are plotted in a solid line, the *allegro* values are plotted in a dotted line, and the *moderato* values are plotted with a dashed line. The amplitude of the different tempi is plotted in figure 9.32.

The clarinet sounds here are generally not the same as the clarinet sounds in section 9.5, which consist more of tenuto executions.

9.7.1. Spectral Envelope Parameters

The spectral envelope parameters for the clarinet with different tempi are shown in figure 9.31. The top left plot is brightness, the top right plot is the tristimulus, the bottom left plot is the odd value and the bottom right plot is irregularity. In all plots the solid line denotes the complete clarinet values, the dotted line the *allegro* values and the dashdotted line the *moderato* values.

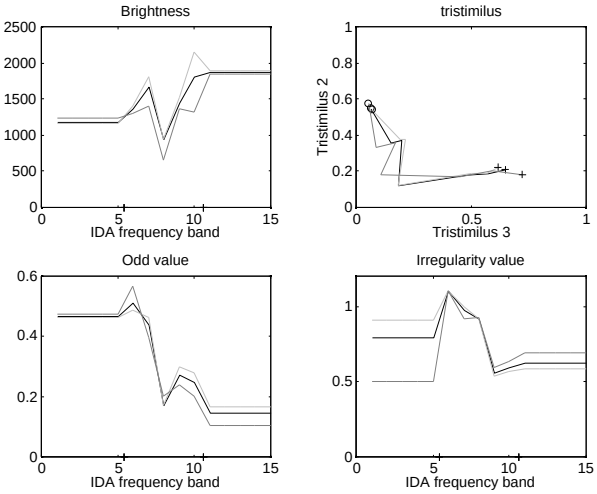


Figure 9.31. Spectral envelope parameters for the clarinet with different tempi. Total (solid), *allegro* (dotted) and *moderato* (dashdotted).

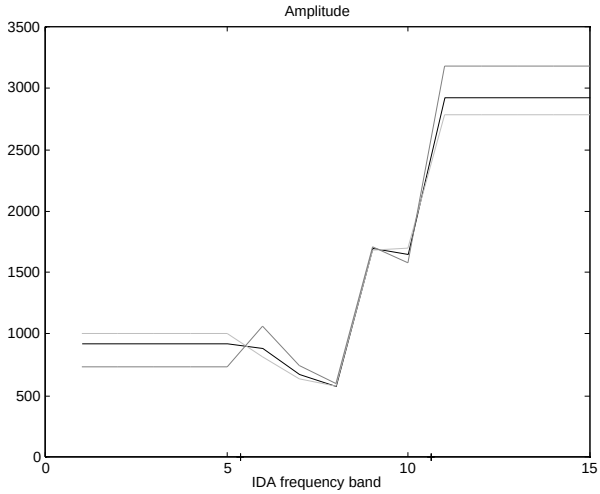


Figure 9.32. Amplitude for the clarinet with different tempi. Total (solid), *allegro* (dotted) and *moderato* (dashdotted).

There seems to be no significant differences in the spectral envelope for the different tempi. Brightness is perhaps a little higher for the high notes of the *allegro* execution.

Brightness is rather constant for the full playing range of the clarinet. The tristimulus first heads towards the fundamental, but then deviates towards the midrange in the upper half of the playing range. This might be explained by the change of register.

The odd value, which is calculated from the third partial, is falling with the fundamental frequency. This is because of the rising fundamental amplitude.

The irregularity value starts above 1. There is generally a very high irregularity because of the weak even partials.

The amplitude curves are very similar for the different tempi. The low fundamental frequency amplitudes decrease slightly, whereas the high fundamental frequency amplitudes increase from below 1000 to above 3000. This is the case for all tempi. Only the edge (lowest and highest fundamental frequencies) values are different for the different tempi.

9.7.2. Frequency Parameters

Inharmonicity is generally very close to zero for the clarinet and it is not plotted here. There is not very much irregularity in the inharmonicity, even for the high error-prone notes. This may be because of the high number of notes used here, by the combination of *piano* and *forte* executions. The clarinet is not expected to have any inharmonicity, and none is found here. It is also the only instrument with absolutely no inharmonicity in figure 9.10.

9.7.3. Envelope Parameters

The attack and release envelope parameters for the different tempi of the clarinet are shown in figure 9.33. The left plots are the attack values, the right plot are the release values. Top plots are the times, middle plots are the percents and bottom plots are the curve form values. The sustain curve form (top) and time (bottom) are shown in figure 9.34.

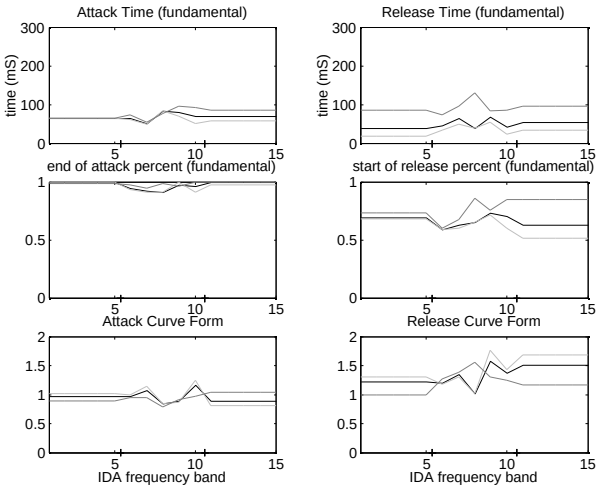


Figure 9.33. Attack and release envelope parameters for the different tempi of the clarinet. Total (solid), *allegro* (dotted) and *moderato* (dashdotted).

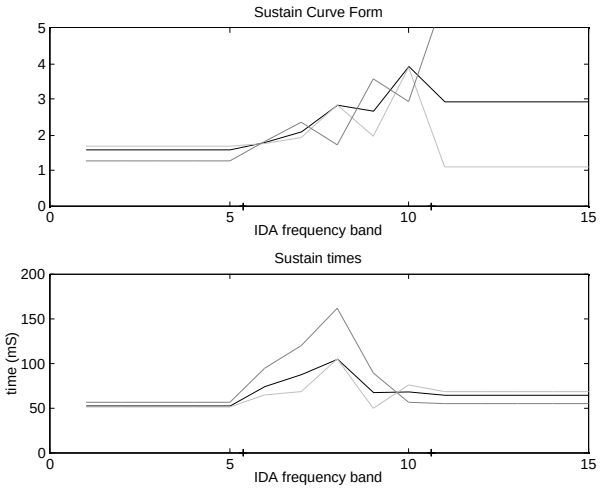


Figure 9.34. Sustain curve form (top) and times (bottom) for the different tempi of the clarinet. Total (solid), *allegro* (dotted) and *moderato* (dashdotted).

The envelope parameters are a priori the parameters that change most with tempo.

The low note attack parameters are not influenced by the different tempi but the high *moderato* notes have a higher attack time. The release times are higher for the *moderato* clarinet, which also has a slightly higher percent value for the release. The sustain times are of course longer for the *moderato* than for the *allegro* sound, although for the highest fundamental frequency this situation is inverted. This is compensated for by the longer release times. Since the percents of the moderate sounds also rise with the fundamental frequency, it would seem that the longer release and shorter sustain times for the highest notes are an effect of the estimation of the parameters. The decay is more exponential for the high fundamental frequencies.

9.7.4. Noise Parameters

The noise parameters for the different tempi of the clarinet are shown in figure 9.35. The left plots are the shimmer values and the right plots the jitter values. The top plots are

the standard deviation, the middle plots the filter coefficient values and the bottom plots the correlation values.

There is not much significant change in the noise values for the different tempi. The main observation is again the similarity of the curves for the different tempi, which shows the success of the parameter estimation.

The shimmer and jitter std for the *moderato* sounds seems slightly lower than the *allegro* values.

The shimmer std seems stable at around 0.2, and the jitter std at around 0.02.

Both the shimmer and the jitter noise coefficient fall with the fundamental frequency.

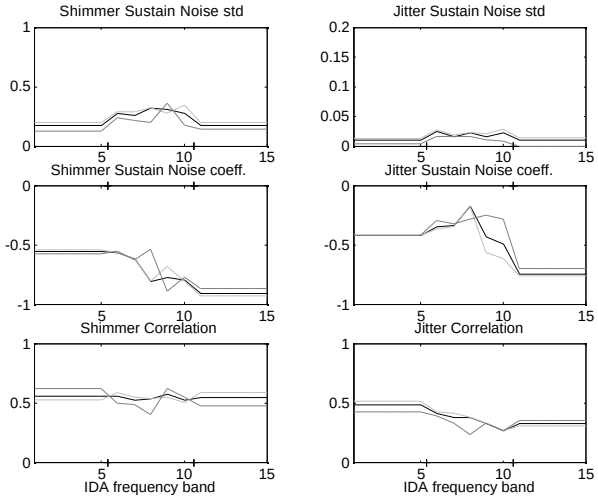


Figure 9.35. Noise parameters for the different tempi of the clarinet. Total (solid), *allegro* (dotted) and moderato (dashdotted).

Shimmer correlation is constant, whereas jitter correlation falls slightly with the fundamental frequency.

9.7.5. Tempo Conclusions

The envelope times are the parameters that change the most with the change in tempo. The attack times do not change significantly, but the sustain and release times do. The sum of the sustain and release times is always larger for the moderato than for the *allegro* executions. The spectral envelope does not change with the tempo. There is more jitter and shimmer for the *allegro* executions than for the *moderato* executions.

9.8. Style

The influence of the style of execution on the timbre attributes is analyzed here. Three different styles are analyzed for the cello: *staccato*, *spiccato* and *legato*. The loudness of all executions is *mezzo forte*, and the tempo is *moderato* for the *legato*, and *allegro* for the *staccato* and *spiccato*. The difference in tempo is unfortunate and does not facilitate the identification of pertinent attributes in the style dimension.

In the following figures, the combination of all cello sounds is plotted with a solid line, the *staccato* executions with a dotted line, the *spiccato* executions with a dashdotted line

and the *legato* executions with a dashed line. The amplitudes of the different styles of the cello are shown in figure 9.37.

9.8.1. Spectral Envelope Parameters

The spectral envelope parameters for the different styles of the cello are shown in figure 9.36. The top left plot is brightness, the top right plot is tristimulus, the bottom left plot is the odd value and the bottom right plot is irregularity. The solid lines are the values for all sounds, the dotted line for the *staccato* sounds, the dashdotted line for the *spiccato* sounds and the dashed lines for the *legato* sounds.

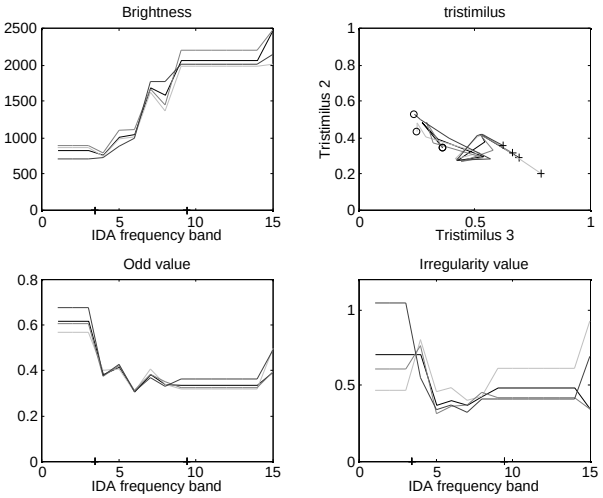


Figure 9.36. Spectral envelope parameters for the different styles of the cello. Complete cello set (solid), *staccato* (dotted), *spiccato* (dashdotted) and dashed (*legato*).

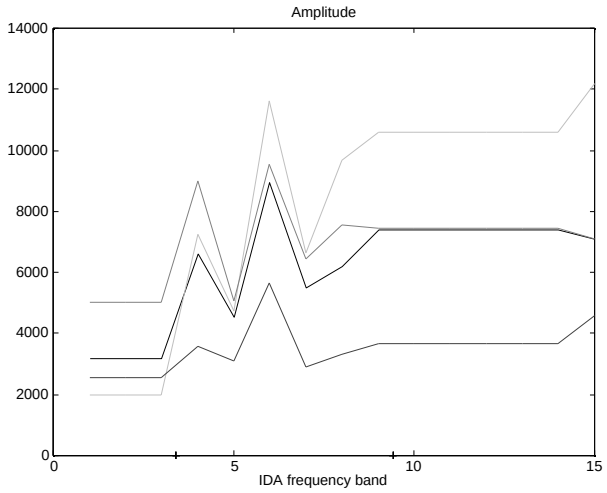


Figure 9.37. Amplitudes for the different styles of the cello. Complete cello set (solid), *staccato* (dotted), *spiccato* (dashdotted) and dashed (*legato*).

Brightness, odd and irregularity values are not influenced very much by the different styles.

Brightness is rising with the fundamental frequency from around 700 Hz for the low notes to around 2 kHz for the high notes.

The tristimulus has a funny loop in the middle for all three styles. Here, first the T2 is increasing, then the T1, then the T3, and finally the T2 is increasing again. This may be an indication of a low resonance, in which first the midrange partials, and then the fundamental is placed.

The odd value is decreasing, but again, this is more because of decreasing partial index brightness, than because of a change in the odd value.

Irregularity starts at a rather high value for the lowest fundamental frequencies, decreasing fast to a stable value of about 0.4 for all styles.

The amplitudes have more differences than would be expected from the relatively homogenous spectral envelope parameters. The *legato* has a much lower amplitude than the other styles.

9.8.2. Frequency Parameters

Inharmonicity is very low for all the cello notes, indicating good initial frequency estimation for this instrument. If anything, inharmonicity is falling slightly with the fundamental frequency for all styles.

9.8.3. Envelope Parameters

The attack and release envelope parameters for the cello are shown in figure 9.38. The left three plots are the attack values and the right three plots the release values.

Top plots are the envelope times, middle plots the percents and bottom plots the curve form values. Complete cello set values are plotted with a solid line, *staccato* values with a dotted line, *spiccato* values with a dashdotted line and *legato* values with a dashed line.

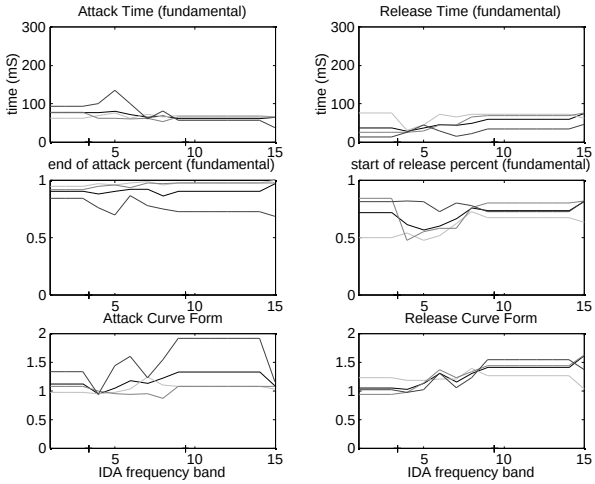


Figure 9.38. Envelope parameters for the different styles of the cello. Complete cello set (solid), *staccato* (dotted), *spiccato* (dashdotted) and dashed (*legato*).

The *legato* has a longer attack time for the low notes, but a shorter attack time for the high notes, and a shorter release time for all notes. This is obviously dependent on the segmentation of the *legato* sounds, if the separation is made close to the attack, then the attack times are short, otherwise the release times are short.

In the *legato* style, the notes are almost glued together, and the total attack and release time should be shorter than the *staccato* or *spiccato* times. This is probably the reason for the short release times for the *legato*.

The *legato* sound also has a smaller attack percent value, indicating a softer attack, with no clear peak at the end of the attack. This is also corroborated with the larger attack curve form values for the *legato* sound, indicating a more logarithmic form in the attacks of the

legato sounds. The *legato* release percents are also higher than the percents for the other executions. The release curve form values are close to one for all executions and rising slightly with the fundamental frequency. The main difference between the *staccato* and the *spiccato* is that the *staccato* has longer release times.

The sustain curve form (top and times (bottom) for the different executions of the cello are shown in figure 9.39 and the start (top) and end (bottom) times are shown in figure 9.40. The decay curve form values are more noisy for the longer *legato* sounds, which have perfectly sustained envelopes.

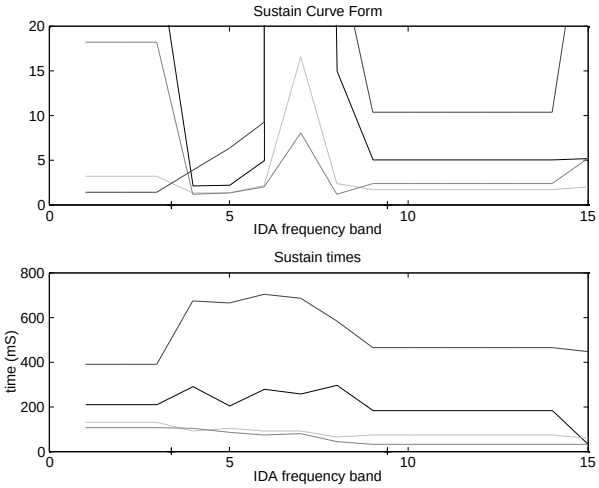


Figure 9.39. Sustain curve form (top) and times (bottom) for the different styles of the cello. Complete cello set (solid), *staccato* (dotted), *spiccato* (dashdotted) and dashed (*legato*).

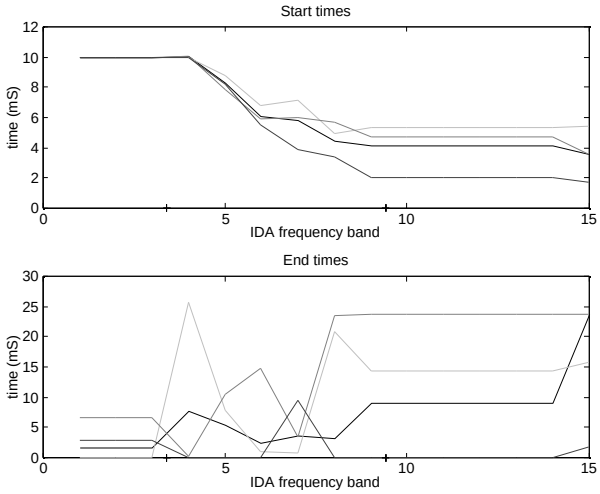


Figure 9.40. Start (top) and end (bottom) times for the different styles of the cello. Complete cello set (solid), *staccato* (dotted), *spiccato* (dashdotted) and dashed (*legato*).

No difference has been found so far between the *spiccato* and the *staccato* sounds. When listening to the sounds, it seems that the largest difference is in the delay between the sounds. This might be lost in the segmentation of the sounds, but it can be deduced from the sustain times in figure 9.39. Here the *spiccato* sounds are shorter than the *staccato* sounds. Since the tempi are equivalent for the two executions, the conclusion is that *spiccato* sounds have more silence between the sounds.

The start and end times for the fundamental and the different styles are shown in figure 9.40. Both the start and end times are very short, indicating that segmentation is done close to the attack and release for all sounds.

In conclusion, the style change on the envelope parameters is hard to detect for these sounds, since they do not have the same tempo. The *legato* has higher attack times, lower attack percents and higher attack curve form values. This is all an indication of softer attack. The *legato* has a shorter release time and higher release percents, which are

indications that the sound has been cut off, either by the segmentation, or by the execution, preparing for the next note. The *staccato* and *spiccato* styles are mainly differentiated by the pause between the notes, *spiccato* having a longer pause than *staccato*, but this is masked by the segmentation of the sounds.

9.8.4. Noise Parameters

The noise parameters for the different executions of the cello are shown in figure 9.41. The left plots are the shimmer values and the right plots the jitter values. The top plots are the standard deviation, the middle plots the filter coefficients and the bottom plots the correlation. Complete cello set values are plotted with a solid line, *staccato* values with a dotted line, *spiccato* values with a dashdotted line and *legato* values with a dashed line.

Few significant differences among the styles have been found in the noise parameters.

Shimmer standard deviation is rather stable for all styles at around 0.2. Jitter standard deviation is very small, with the exception of the low *staccato* sounds, which may be caused by bad analysis.

The *legato* shimmer filter coefficients are much lower than the other filter coefficients. The relatively longer *legato* sounds and consequently worse envelope model may explain this.

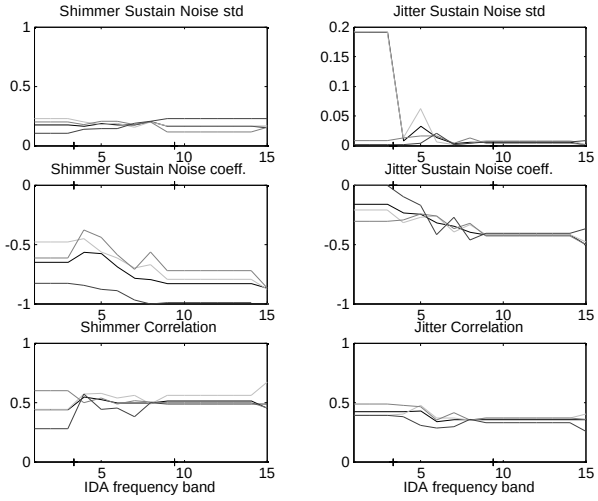


Figure 9.41. Noise parameters for the different styles of the cello.

The jitter filter coefficients are very close to zero, indicating a band-pass noise. Correlation is close to 0.5 for both shimmer and jitter for all sounds and styles.

9.8.5. Style Conclusions

The conclusions from the analysis of the cello sounds with different styles are made difficult because of the different tempi of the different styles and because of the uncertainty caused by segmentation.

Nonetheless, the *legato* was generally found to have a longer attack and short release time, and a more rounded attack. It also has shorter release and higher release percents.

The style seems to influence only the envelope parameters, but a difference in shimmer filter coefficient was detected for the *legato* sounds. This is explained by the worse fit on long sounds of the simple envelope model, which cannot model tremolo, or other voluntary amplitude variations. No changes were detected in the spectral envelope values for the different styles of the cello. The difference between the *staccato* and *spiccato* styles is mostly related to the pause between the sounds.

9.9. Sound Recreation from IDA Parameters

An MDA parameter set for a given fundamental frequency is extracted from the IDA parameter set, by finding the corresponding IDA frequency band, and copying the parameters from this band into the MDA. The HLA parameters and additive parameters can then be created, and finally a sound can be synthesized.

Morphing between different IDA classes is done by multiplying each IDA with a coefficient, and ensuring that the sum of all coefficients equals one.

The quality of the IDA sound synthesis is generally very close to the MDA quality. No significant difference in quality has been found, so the conclusions from the MDA in Chapter 8 are also valid here. No improvement or deterioration has been found from the summation of many MDA parameters.

9.10. Conclusions

The IDA parameter set is helpful in analyzing the evolution of timbre attributes across the playing range of an instrument. Some timbre attribute evolutions are common for many instruments, whereas others have individual evolutions for each instrument, which can sometimes be explained by the characteristics of the instrument. The analysis of the IDA parameter evolution across playing range, loudness or other classes, is helpful in understanding which timbre attributes are responsible for what sound quality change.

The partial index brightness is decreasing with fundamental frequency, giving most of the amplitude to the fundamental for the highest fundamental frequencies. The frequency brightness is increasing with the fundamental frequency for most instruments. One exception is the violin, which has a fairly constant brightness. The attack time and to a lesser degree the release times decrease with the fundamental frequency. Again the exception is the violin, which has a fairly constant attack time. Noise is not very influenced

by the fundamental frequency, although the shimmer standard deviation often rises with the fundamental frequency.

A change in loudness influences only the spectral envelope parameters, notably brightness and amplitude, which both increases with loudness. A decrease in tempo modifies the envelope time by increasing the sustain and release times. The shimmer filter coefficient decreases with a decrease in tempo, indicating a low-frequency rumbling irregularity, which is probably caused by the poor envelope model. The change in style also modifies the envelope parameters most, the *legato* having softer attack and shorter releases. The difference between *staccato* and *spiccato* is related to the length of the silence between sounds.

The IDA parameters generally recreate the identity of the instrument in resynthesis but the sound quality is the same as for the MDA parameters. Although it is believed that the IDA parameters are sufficient for a good resynthesis of a musical sound, more work remains before this can be achieved.

Chapter Ten

10. Timbre Modifications

One of the exciting things to do with the timbre models is to modify the parameters and listen to the effect on the sound. Since the timbre model attributes are well understood, the modifications on the timbre are intuitive and the effects that are sought for are easily implemented. The modifications can be either of the expression parameters, such as the pitch, or the identity of the sound. All timbre models can be modified, or used as templates in the modification of another model, but this chapter focuses on the modification of the additive parameters. Furthermore, the concatenation of two sounds is discussed. The modifications presented in this chapter can be used to ‘play’ the different timbre models.

10.1. Introduction

This chapter discusses the modifications of the parameters of the timbre models. The modifications can involve either an expression, such as the pitch, or the identity of the sound. The expression modifications must be possible in real-time, ‘on the fly’, and this is also highly desirable for the other modifications, although no special attention has been put into this problem.

Most people who work in the analysis/synthesis of sounds do expressive or timbre manipulations. [Bode 1984] offers a non-exhaustive review of early sound modifications. [Allen 1977] modified the sound in the fourier domain, [Quatieri *et al.* 1986] made speech transformations in the additive domain. Other synthesis techniques with good possibilities for timbre manipulation are the granular synthesis [Roads 1988] and the physical modeling [Jaffe *et al.* 1983]. [Lent 1989] proposes a method for efficient pitch shifting of sounds. [Fitz *et al.* 1996] made timbre manipulation using additive parameters, and [Tellman *et al.* 1995] used the same parameters to do timbre morphing. [Rovan *et al.* 1997] made expressive changes in the additive domain and [Arcos *et al.* 1997] uses case-based reasoning system to generate expressive musical performance with SMS [Serra *et al.* 1990]. The diphone program [Rodet *et al.* 1997] is also used for the manipulation and concatenation of additive or other parameters.

The advantage of this work is to have available a complete timbre model, as for instance the HLA model presented in Chapter 6, or the MDA model presented in Chapter 8, into which the additive parameters are to be shaped.

Several types of modifications can be made, first, there are inter-model modifications, where two parameter sets from the same model are combined, and then there are extra-model modifications where the parameters from one model are used to modify the parameters of the other model. Furthermore, there is the timbre morph, where two sounds are transformed into another sound, which is somewhere intermediate in the timbre space. Individual timbre attribute modification, where one or a few attributes are modified, is also possible.

The modifications of the HLA or MDA parameters, or between the HLA and the MDA parameters are relatively straightforward, therefore this chapter focuses on the modification of the additive parameters with a HLA template.

The modifications of the expressive parameters pitch, loudness and duration are discussed in section 10.2 and the inter model modifications are discussed in section 10.3. Section 10.4 talks about the concatenation of two sounds and the modification method developed in this work of the additive parameters is presented in section 10.5. The quality of the resynthesis is discussed in section 10.6 and the chapter ends with a conclusion in section 10.7.

10.2. Pitch, Loudness and Duration

Here the most important parameters of the timbre are modified. These modifications can be made on any type of model, and independent of modifications of other parameters. Generally, the modifications can be made real-time, thus enhancing the vivacity of an instrument.

10.2.1. Pitch

The pitch of a sound is the perceived fundamental frequency of a sound. The modification of the pitch is done here by modifying the mean frequencies of the individual partials. Therefore, changing the pitch from sound[a] to sound[b] consist of modifying the mean of each partial frequency from the value from sound[a] to the value from sound[b],

$$f_k(t) = f_k^a(t) \frac{\bar{f}_k^b}{\bar{f}_k^a} \qquad \textbf{(10.1)}$$

where $f_k^a(t)$ is the original time-varying frequency of partial *k*. This means that only the mean frequency of each partial is changed. This is true for the additive and the HLA models. For the MDA model, where the frequency is modeled by the fundamental frequency and inharmonicity index, these two values are changed in order to change the pitch. When an intermediate value is wanted, the frequencies are modified by a ratio times parameter[a] plus another ratio times parameter[b],

$$f_k(t) = f_k^a(t) \frac{(1-r)\ \bar{f}_k^a + r\ \bar{f}_k^b}{\bar{f}_k^a} \qquad \textbf{(10.2)}$$

where $0 \le r \le 1$. In the case of the HLA model, the frequencies have no time index, which can thus be omitted. If the individual partial frequencies are not available, all frequencies are changed by the same ratio.

The aforementioned modifications can be stored in all the models, but the vibrato effect can only be stored in the additive model. Vibrato consists of multiplying an offset low-frequency waveform *vib(t)* (typically a sinusoidal) to the frequency of the partials,

$$f_k(t) = f_k(t)\ (1 + vib(t)) \qquad \textbf{(10.3)}$$

Typical values of the vibrato are a few percents, and often the vibrato is delayed slightly, but the values of these parameters are chosen at performance time, and not further discussed here.

10.2.2. Loudness

Loudness is the perceived intensity of a sound. Loudness is modified by changing the maximum of the amplitude of the partials. The modifications of the amplitudes are made in a manner similar to the pitch modifications. The transformation from sound[a] to sound[b] is done by multiplying the time-varying amplitudes $a_k^a(t)$ by the maximum of the target amplitude, and dividing by the maximum of the original amplitude for each partial *k*,

$$a_k(t) = a_k^a(t) \ \frac{\max(a_k^b(t))}{\max(a_k^a(t))} \qquad \textbf{(10.4)}$$

In the case of morphing between two sounds, the amplitude is,

$$a_k(t) = a_k^a(t) \ \frac{(1-r) \ \max(a_k^a(t)) + r \ \max(a_k^b(t))}{\max(a_k^a(t))} \qquad \textbf{(10.5)}$$

where $0 \le r \le 1$. Again, in the case of the HLA model, the frequencies have no time index, which can thus be omitted. If the individual amplitudes of the target sound are not available, all amplitudes are changed by the same ratio. The MDA model is modified by setting the spectral envelope parameters.

The slow oscillating of the amplitudes is called tremolo, and tremolo can only be stored or added to the additive model. Tremolo is created by adding a low-frequency waveform *trem(t)* to the amplitude of the partials,

$$a_k(t) = a_k(t) \ (1 + trem(t)) \qquad \textbf{(10.6)}$$

By multiplying with the original amplitudes it is ensured that there is no tremolo in the silence, and that the tremolo decreases gracefully when the amplitude decreases. Normal tremolo values are between 10% and 50%. A tremolo ten times the vibrato seems to give about the same perceptive effect.

10.2.3. Duration

The duration is the perceived length of the sound. The duration is here modified by changing the sustain length of each partial,

$$t_{sustain,k}^b = t_{sustain,k}^b + t_{\mathrm{mod}} \qquad \textbf{(10.7)}$$

where

$$t_{\mathrm{mod}} = t_{length}^b - t_{length}^a \qquad \textbf{(10.8)}$$

is the difference of length between sound^a and sound^b. If $t_{\mathrm{mod}} < 0$, the new sustain length can become negative. If this happens, it is necessary to decrease the attack and release times, to ensure that the envelope has the right length. It is important that all partial lengths are changed by the same amount, since otherwise the individual release times will not be synchronized, which is very perceivable.

If the absolute value of t_{mod} is large, it might be necessary to either duplicate or cut out parts of the sustain segment, to maintain the same noise frequency magnitude.

The modifications of the sustain length can give rise to artifacts in the sound if it is an attack-decay-release type of sound. In that case, it is necessary to change the decay slope to accommodate for the change in length. In for instance the piano, augmenting the length without changing the slope gives the sound an unnatural strength in the end of the decay. The slope is defined in the HLA model by the percent values at the end of attack and the start of release and the curve form. The sound can very well be of sustain type even though the start of release value is much lower than the end of attack value, if the curve form is 'bending over' in the end. Nevertheless, the curve form is ignored, and the decay is simplified into a linear form, where the value at the start of release is changed, if it is smaller than the end of release value,

$$v_{sor} = v^a_{eoa} + (v^a_{sor} - v^a_{eoa}) \frac{t^b_{sustain}}{t^a_{sustain}} \qquad \textbf{(10.9)}$$

Obviously, this value is truncated at zero, since the amplitude can not be negative. Although the linear shape is a simplification of the real decay form, the modification of the start of release value according to a linear model seems to correct the sustain length problem.

10.2.4. Number of Partials

The number of partials is a fairly simple attribute. Although it may be important when the higher partials contain much energy, in general, the number of partials is not very crucial. It can be important, nonetheless, to add or remove partials, for instance when combining a high flute sound with a low piano sound.

In general terms, if the goal is to transform sound^a into sound^b, and sound^a has N^a partials, and sound^b has N^b partials, then the resulting sound should have N^b partials.

If $N^a > N^b$ then the N^b+1 to N^a partials are removed from the sound^a. This is easily done, whether sound^a is modeled by additive parameters, or HLA parameters. The MDA and

IDA models don't include the partial number, so the question doesn't arise for these models.

If $N^b > N^a$, partials must be added to sound^a. If sound^a is modeled by the HLA, then the corresponding MDA parameters are calculated, and a new HLA with N^b parameters is created. The N^a+1 to N^b partials from the new HLA are then copied to the same partials in the original HLA. If sound^a is defined by additive parameters, then two methods can be used to create the N^a+1 to N^b partials. The first consists of creating the corresponding HLA and MDA, and a new HLA with N^b partials, and finally creating N^a+1 to N^b synthetic partials which are added to the additive parameters of sound^a. The second method, which has been adopted here, consists of copying the N^a partial to all the missing partials, and modify only the amplitude and frequency of these partials, so they correspond to the upper partials from the synthetic HLA of the same sound, with N^b partials.

If a sound is created in between sound^a and sound^b, it can be supposed that the number of partials is an intermediate value, and the number of partials of the resulting sound is,

$$N = r \quad N^a + (1 - r) \quad N^b \qquad \textbf{(10.10)}$$

where $0 \le r \le 1$.

10.3. Inter-Model Modifications

In this section the modification of sound^a into sound^b is discussed. Both sounds are supposed to be defined by the same model, be it the additive, the HLA, or the MDA model.

The additive parameter model consists of *k* partials with time-varying amplitude and frequency.

The HLA model is presented in Chapter 6. It consists of the spectral envelope, the mean frequencies, the envelope times, percents and curve form values, the shimmer and jitter values. Most parameters have values for each segment of the envelope, and all values are individual for each partial.

The MDA is presented in Chapter 8. It consists of the same parameter types, but the spectral envelope is modeled by brightness, tristimulus 1 and 2, the odd value, irregularity and maximum amplitude. The mean frequencies are modeled by the fundamental value and inharmonicity, and all other parameters are modeled by the fundamental value and the value of an exponential, which models the evolution across the partial index.

The modification of the additive parameters necessitates the normalization of the length. This is done by changing the full length of each partial using the method exposed in paragraph 10.5.3.1. Furthermore, the number of partials must be equal for both sounds. This is done by the method presented in 10.2.4. The amplitudes and frequencies of sound^a and sound^b are then morphed by the following formulas,

$$a_k(t) = r \;\; a_k^b(t) + (1-r) \;\; a_k^a(t) \qquad \textbf{(10.11)}$$

$$f_k(t) = r \;\; f_k^b(t) + (1-r) \;\; f_k^a(t) \qquad \textbf{(10.12)}$$

More than two sounds can of course be used to create the output parameters.

This method suffers from the lack of temporal cues, and an improved method of the modification of the additive parameters is introduced in section 10.5.

If the HLA^a is to be modified partly into HLA^b, the number of partials must first be normalized, as explained in paragraph 10.2.4. Then the HLA sets can be combined easily, in fact then,

$$HLA = r \;\; HLA^b + (1-r) \;\; HLA^a \qquad \textbf{(10.13)}$$

Here, the spectral envelope features might not be well interpolated.

The MDA models are combined, or morphed, ‘straight out of the box’,

$$MDA = r \;\; MDA^b + (1-r) \;\; MDA^a \qquad \textbf{(10.14)}$$

Care must be taken when morphing the filter coefficient values for shimmer and jitter if the fundamental frequency of the sounds are very different, since the filter coefficient values are dependent on the sampling rate of the additive parameters, which is equal to the fundamental frequency, f_0. (cf. the data reduction in section 4.3.5 in Chapter 4).

In conclusion, the inter model modifications are relatively straight forward, but the averaging method used can remove important information, such as the noise, and the interpolation of perceptually important features may not be handled properly in the good quality additive or HLA models.

10.4. Concatenation

When playing two sounds one after the other, the sounds need to be concatenated. If the sounds are played with a silence between them, the sounds can be created individually, and then concatenated with the suitable silence in between. However, if the second sound is

interrupting the first, care must be taken so that the two sounds sound right. The concatenation can define important timbre cues, such as the style of the execution.

The concatenation involves the times of the envelope. The start and end segments are minimized to the smallest length that permits a common start and end of the sound. This restitutes the effective duration of the sound. The original start and end segments are used only if there is a pause between the sounds.

Two types of concatenation are possible, the superposition of the new sound parameters on the old, and the replacement of the old by the new sound parameters. Superposition is used for example when changing the string in the violin and replacement is used when playing a new note on the same string.

Concatenation is here discussed on the additive parameters, using the HLA model as a template for the sounds.

10.4.1. Superposition

The superposition of sound^{b} on or after sound^{a} is done simply by adding the additive parameters of sound^{b} at time t to the additive parameters of sound^{a}. Timing is ensured by setting the time zero of sound^{b} as the time of the earliest start of attack.

Superposition is also used when playing chords. One problem with the superposition of additive parameters is the large number of partials that results from the superposition of several sounds. Cleaning out masked [Small 1959] partials could potentially reduce the number of partials resulting from superposition.

10.4.2. Replacement

Replacement is the normal mode when playing a melody with the additive parameters. Here, the amplitude and frequency of the partials of sound^{a} are transformed into the values of sound^{b}. The time of the transformation, t_p, can be set, and it defines the length of the portamento of the transition.

Assuming that the end of sound^{a}, t_e^a, and the start of sound^{b}, t_s^b, are available. t_s^b is defined as the earliest attack time, and the partials of sound^{b} prior to this time are not used. The parameters of sound^{b} are inserted at time $t+t_p$, and the amplitudes and frequencies of sound^{a} at time t are modified into the amplitudes and frequencies of sound^{b} at time t_s^b. If the sound^{a} is longer than $t+t_p$, the remaining part is not used.

The frequencies at the portamento segment of the output sound are the sum of the two frequencies, multiplied by two curves that ensure that the frequencies change from those of sound[a] to those of sound[b].

$$f_k(t) = f_k^a(t)\ \ curve(0, \frac{f_0^b}{f_0^a}, t_p) + f_k^b(t)\ \ curve(\frac{f_0^a}{f_0^a}, 0, t_p) \qquad \textbf{(10.15)}$$

where f_0 is the fundamental frequency. The function *curve(a,b,t)* makes a curve from value *a* to value *b* with length *t*. The curve form is not specified here, but it could instead be specified at performance time, preferably using some real-time sensor.

The replacement can also be used to play a melody with the same additive parameter set, but if the pitch is changed by more than half an octave, the sound generally loses realism, and it needs to be modified as explained in the next section.

A simple modification would be to change only the brightness; the resulting brightness should be (cf. The IDA analysis in Chapter 9),

$$Br = Br_0 / \left(\frac{\log(f_0^b)}{\log(f_0^a)}\right)^2 \qquad \textbf{(10.16)}$$

where Br_0 is original brightness. This change means brightness decreases when the pitch is increased. Other parameters, which seem to have a constant evolution across pitch, independent of the instrument, are tristimulus, odd, amplitude and attack time. These parameters must be changed if a natural sound is wanted across a large pitch or intensity range. More information on the evolution of the timbre attributes as a function of pitch can be found in Chapter 9.

10.5. Additive Modifications

In this section, the transformation of the additive parameters of sound[a] into sound[b] is discussed. The sound[b] is defined by the HLA parameters. The sound[a] HLA parameters are also available. The HLA parameters are indicated with a hat on the parameters. The transformation is done in several steps: first the spectral envelope is transformed, then the pitch, then the amplitude envelope is transformed in several steps, and finally the shimmer and jitter are transformed. All of these steps can be done individually, if only some of the timbre attributes should be changed. The final modifications give a sound that is very close to sound[b]. In the following, the transformation of a piano sound (a) into a trumpet sound (b) is illustrated.

10.5.1. Spectral Envelope

The amplitudes are modified in the same manner as explained in paragraph 10.2.2. The time-varying amplitude of each partial is multiplied by the static spectral envelope of sound[b] and divided by the static spectral envelope of sound[a],

$$a_k(t) = a_k^a(t) \ \frac{\hat{a}_k^b}{\hat{a}_k^a} \qquad \textbf{(10.17)}$$

The hat on the parameters indicates here that they are derived from the HLA model. $\hat{a}_k$ is the spectral envelope value at partial *k* (the maximum amplitude of the partial *k*).

The spectral envelope of the original and the modified additive parameters of the piano sound are shown in figure 10.1. The trumpet spectral envelope is not shown, since it is identical to the modified piano spectral envelope.

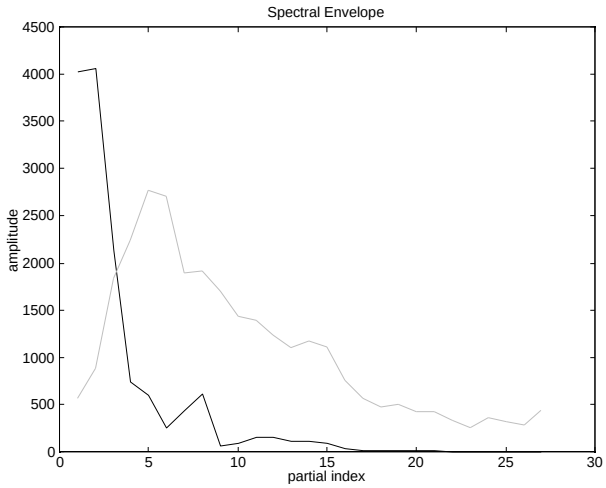


Figure 10.1. Spectral Envelope of the original (solid) and modified (dotted) piano.

The modification of the spectral envelope is a powerful modification, which changes the sound substantially. Nonetheless, the identity of the sound is not transformed just by modifying the spectral envelope, since the same instrument can have very different spectral envelopes in different playing ranges.

This is the first step in transforming sound[a] into sound[b]. The next step is to change the frequencies of the output additive parameters. This changes the pitch of the resulting sound.

10.5.2. Frequency

The frequencies are modified in the same manner as explained in paragraph 10.2.1. The frequency of each partial is multiplied by the new frequency envelope and divided by the old frequency envelope,

$$f_k(t) = f_k^a(t) \ \frac{\hat{f}_k^b}{\hat{f}_k^a} \qquad \textbf{(10.18)}$$

where the HLA frequency $\hat{f}_k^a$ in this case is the mean frequency of partial *k* in the sustain region.

The original and transformed frequencies of the piano sound are shown in figure 10.2. The original piano frequencies, divided by the partial index, constitute the solid line and the modified piano frequencies, which are identical to the trumpet frequencies, constitute the dotted line. It is clear that the original piano has a greater inharmonicity index than the trumpet. The fundamental frequencies are very close, so the sound is not changed very much by this modification. The trumpet has a slightly higher fundamental frequency than the piano, but no partial frequency stretching. The highest frequencies of the trumpet are probably misjudged due to noise.

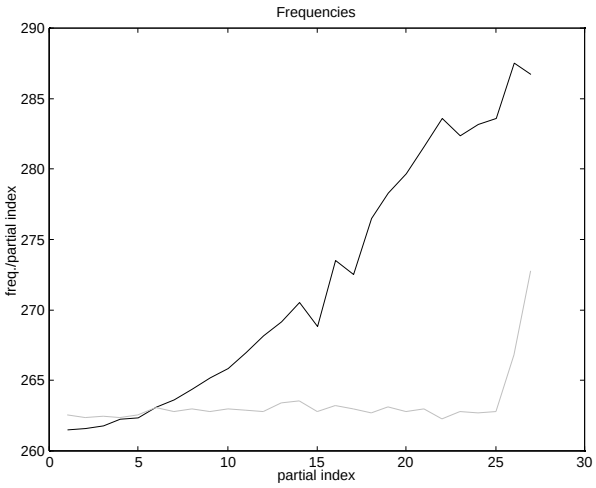


Figure 10.2. Mean frequencies, divided by the partial index, of the original (solid) and modified (dotted) piano.

This is the second step in the transformation of the piano sound into a trumpet sound. The additive parameters now have the same mean frequencies and maximum amplitudes as the resulting sound. The next steps involve the modifications of the amplitude envelope and noise parameters.

10.5.3. Envelope

The envelope is an important timbre attribute. The envelope model is presented in Chapter 5. It is defined for each partial in the HLA model as the times, percents and curve forms of an attack-sustain-release or attack-decay-release model, which also includes a start and an end segment. The envelopes of the partials of the piano sound[a] is transformed into the envelope form of the trumpet sound[b] by first modifying the envelope times, then changing the envelope percents, and finally changing the curve forms.

10.5.3.1 Envelope Times

The envelope times of the sound[a] is changed into the times of sound[b] by adding a linear slope *lincurve* for each segment. The slope has the value zero at the start of the slope and the difference between the new and the old segment length at the end of the slope,

$$t_{s,k}(t) = t^{a}_{s,k}(t) + lincurve(0, \hat{t}^{b}_{s,k} - \hat{t}^{a}_{s,k}, \hat{t}^{a}_{s,k}) \qquad \textbf{(10.19)}$$

The time with a hat and no *t* parameter is the length of the segment from the HLA model. The time with parameter *t* is the time at envelope index *t*. The function lincurve(a,b,t) makes a linear curve from *a* to *b* with length *t*. This accelerates or decelerates the times so it is equal to the new length at the end of the segment. The envelope time modifications are done for the start, attack, sustain, release and end segments for all partials.

The modified envelope of the fundamental is shown in figure 10.3. The modified envelope has first been multiplied so the maximum value is the same as the target value, as explained in paragraph 10.5.1. The target envelope has been created from the clean (noiseless) HLA parameters of the trumpet sound.

The plus signs denote the split points for the original and the wanted envelope.

Already, the envelope has a good resemblance to the target envelope, although the sustain curve form is wrong. These parameters are corrected in the following two paragraphs.

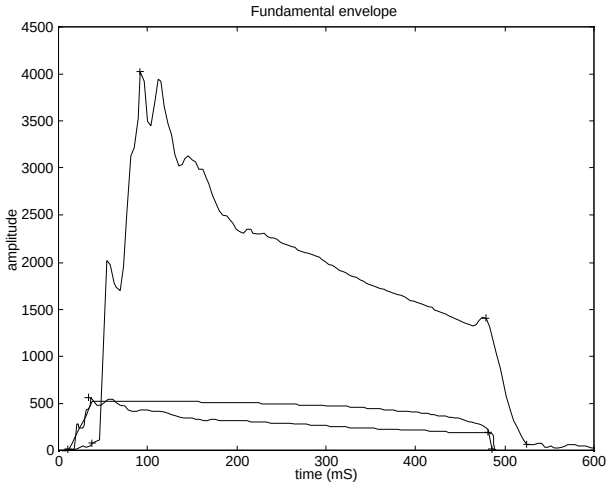


Figure 10.3. Original (top), target, and modified (bottom) fundamental envelope after time modification.

10.5.3.2 Envelope Percents

The next step is to change the envelope percents. The envelope percents are the relative amplitudes at the split points. This modification is done for each percent by multiplying the adjoining segment envelopes with a linear slope, which is 1 at the far ends, and the percent[b] divided by the percent[a] at the split point. The modification of, for instance, the sustain segment of partial *k* is then,

$$a_{s,k}(t) = a^{a}_{s,k}(t) \;\; (lincurve(\frac{\hat{p}^{b}_{eoa,k}}{\hat{p}^{a}_{eoa,k}}, 1, t_{s,k}) \;\; lincurve(1, \frac{\hat{p}^{b}_{sor,k}}{\hat{p}^{a}_{sor,k}}, t_{s,k})) \qquad \textbf{(10.20)}$$

$\hat{p}_{eoa,k}$ is the percent at the end of attack, and $\hat{p}_{sor,k}$ is the percent at the start of release. The envelope percent changes are done for the soa, eoa, sor and eor percents.

The original, target and modified envelopes after the percent modification of the fundamental are shown in figure 10.4. The main difference here is that the percent of the piano in the end of release split point is almost zero, whereas the corresponding trumpet percent is very large. The linear slope value then becomes very large, and the curve can be substantially modified. What happened here is that the characteristic peak at the end of the decay of the piano was amplified, so it is very visible. Since this phenomenon is not included in the HLA model, it is normal and necessary that it is transmitted from the piano to the trumpet sound.

The fundamental partial of the trumpet actually ends in the eor split point. This is not very visible in figure 10.4, but it becomes visible after the resampling is performed in paragraph 10.5.4. Although the envelope doesn't look any closer to the target envelope after this modification, the next step, which is the modification of the curve form, will show that this is the case.

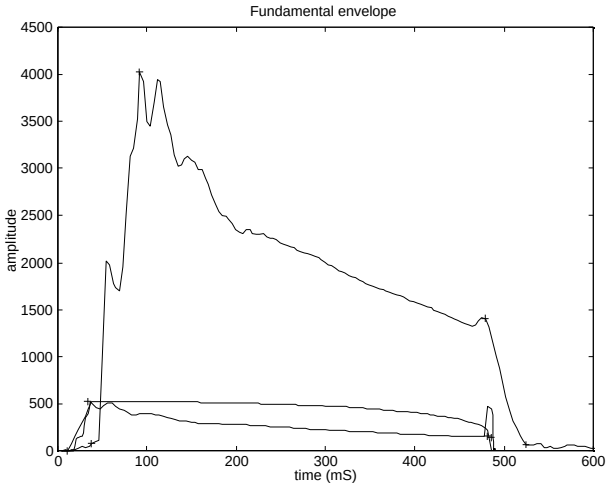


Figure 10.4. Original (top), target, and modified (bottom) fundamental envelope after the percents modification.

10.5.3.3 Envelope Curve Forms

The envelope curve form is the last step in the modification of the amplitude envelopes. The curve form is a simple equation with three parameters, v_1, v_2 and n.

$$curve(t) = v_1 + (v_2 - v_1)\ (1-(1-t)^n)^{\frac{1}{n}} \qquad \textbf{(10.21)}$$

where v_1 and v_2 are the start and end values, which are found by multiplying the percent with the spectral envelope value for the corresponding segment and partial. t is the time index, and n is the curve form.

The modification for each segment from curve^a to curve^b is made by adding curve^b and subtracting curve^a,

$$a_{s,k}(t) = a^a_{s,k}(t) + curve^b(t) - curve^a(t) \qquad \textbf{(10.22)}$$

This can be done, since the length, start and end points are the same for both curves after the timing and percent modifications. The modification of the curve forms is done for the start, attack, sustain, release and end segments.

The envelope of the fundamental of the piano is shown in figure 10.5. It is obvious that the curve form of the piano now is really close to the curve form of the trumpet. The main difference is the peak at the end of the release, which is very characteristic of the piano.

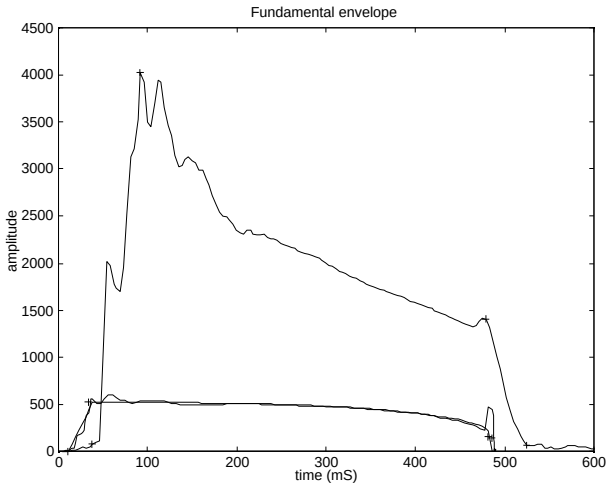


Figure 10.5. Original (top), target, and modified fundamental envelope after the curve form modification.

It is now clear that the previous modifications of the envelope percents actually approached the source envelope to the target envelope even though it sometimes seemed the opposite was happening.

The envelope timing and percent modification are necessary first steps if the curve form modification shall succeed.

The modification of the curve form of the envelope finishes the third step of the transformation of the piano sound[a] to the trumpet sound[b]. The partials now have a close physical resemblance, as can be seen in figure 10.5, but the sound still is different, so the noise attributes are changed next.

10.5.4. Noise Modification

The noise is an important attribute of the timbre, and it is here modeled as the irregularities on the amplitudes (shimmer) and frequencies (jitter) of the partials. The noise model is presented in Chapter 6. The partials need to be resampled after the frequency and envelope modifications, since the noise is calculated on the partials with a sampling rate of the fundamental frequency of the sound. The sampling conversion is done by linear interpolation. Better results may be obtained with better interpolation methods, but so far no sound artifacts have been observed with the simple linear interpolation.

If the fundamental frequency modification has been important, or the segment length changes are important, then it may be necessary to revise the segment length modification

method, since the high frequency content of the noise of, for instance, a very short segment changed into a very long segment, is dilated to be almost non-existent.

10.5.4.1 Shimmer

Shimmer is the noise on the amplitude of the partials. It is modeled in the HLA model with two parameters, the standard deviation and the filter coefficient of a single-tap recursive filter.

The magnitude response of a single-tap recursive filter with one parameter *a* is [Steiglitz 1996],

$$|H(\omega)| = \frac{1}{\sqrt{1 + a^2 + 2a\ \cos(\omega)}} \quad \textbf{(10.23)}$$

The shimmer is first extracted by subtracting the clean sound[b] envelope from the modified sound[a] envelope, and normalizing with the clean envelope,

$$shimmer_{s,k} = \frac{e^b_{s,k}(t) - \hat{e}^b_{s,k}(t)}{\hat{e}^b_{s,k}(t)} \quad \textbf{(10.24)}$$

where $e_{s,k}$ is the amplitude envelope for the partial *k*, and $\hat{e}_{s,k}$ is the clean envelope for the same segment and partial. The shimmer parameters need to be recalculated, instead of extracted from the HLA model, since it has been substantially modified in the previous paragraphs. The filter coefficient is now calculated from the shimmer, and the modified shimmer with the desired frequency magnitude is calculated by multiplying by the new filter magnitude response[b] and dividing by the old filter magnitude response[a] in the frequency domain,

$$shimmer_{s,k} = FFT^{-1}(FFT(shimmer)\ \frac{|H(\omega)^b|}{|H(\omega)^a|}) \quad \textbf{(10.25)}$$

The standard deviation is then calculated from the modified shimmer, and the shimmer is again modified by multiplying by the wanted standard deviation[b] and dividing by the calculated standard deviation[a],

$$shimmer_{s,k} = \frac{\sigma^b}{\sigma^a}\, shimmer_{s,k} \quad \textbf{(10.26)}$$

This creates a new noise with the target standard deviation and filter coefficient values. The noise is reinserted into the clean envelope[b] by multiplying it with the clean envelope[a] and adding it,

$$e^{a}_{s,k} = \hat{e}^{a}_{s,k}(1 + shimmer_{s,k}) \quad \textbf{(10.27)}$$

The sustain shimmer for the fundamental of the piano is seen in figure 10.6 before and after shimmer modification (dotted). The top plot is the frequency domain noise, and the bottom plot is the time domain noise.

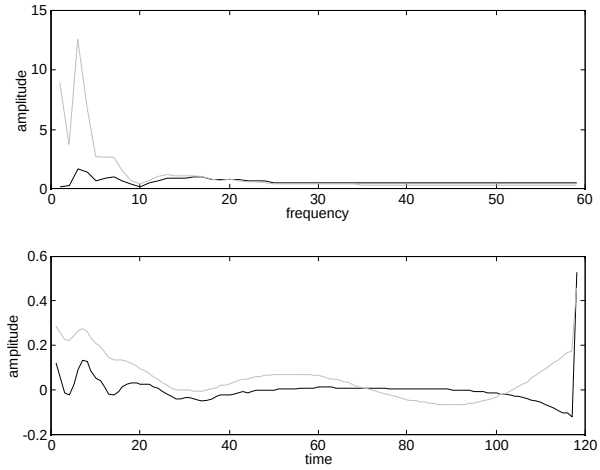


Figure 10.6. Frequency magnitude response (top) and time signal (bottom) for sustain shimmer of the fundamental of the piano, original (solid) and modified (dotted).

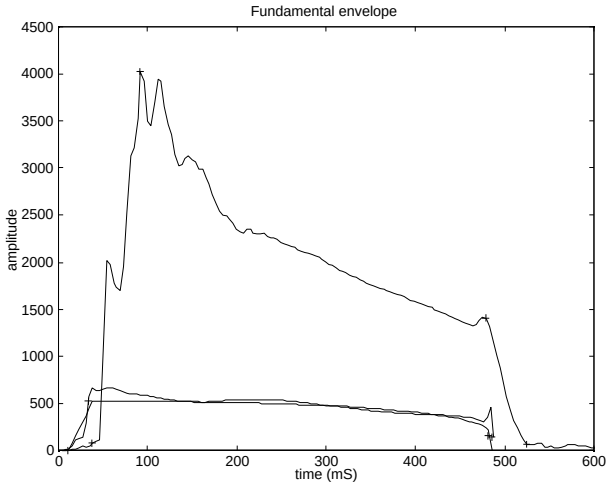


Figure 10.7. Original (top), clean trumpet envelope, and modified piano fundamental envelope after the shimmer modification.

It seems fairly obvious that the trumpet has a more low-pass noise with a higher standard deviation. The standard deviation values are 2% for the piano fundamental, and 9% for the trumpet. The filter coefficient changes from -0.93 for the piano to -0.98 for the trumpet. The resampling has made obvious the large end of release percent for the trumpet sound, which has been cut off before the end of the release.

10.5.4.2 Jitter

Jitter is the irregularities on the frequencies of the partials. The jitter is modeled in the HLA as the shimmer is, with two parameters; the standard deviation and the filter coefficient of a single-tap recursive filter, which has the mean-square frequency magnitude response approximation of the original noise magnitude response. Jitter is calculated as the frequency minus the mean of the frequency, divided by the mean of the frequency for each segment and partial,

$$jitter_{s,k} = \frac{f^{a}_{s,k} - \bar{f}^{a}_{s,k}}{\bar{f}^{a}_{s,k}} \quad \textbf{(10.28)}$$

Jitter is modified in the same manner as the shimmer, first the filter coefficient is calculated from the jitter, then the frequency magnitude response is modified,

$$jitter_{s,k} = FFT^{-1}(FFT(jitter)\ \frac{|H(\omega)^b|}{|H(\omega)^a|}) \qquad \textbf{(10.29)}$$

The standard deviation is then calculated from the modified jitter, and the jitter is again modified by multiplying by the wanted standard deviation[b] and dividing by the calculated standard deviation[a],

$$jitter_{s,k} = \frac{\sigma^b}{\sigma^a}\ jitter_{s,k} \qquad \textbf{(10.30)}$$

Finally, the resulting frequency is,

$$f_{s,k}^a = \hat{f}_{s,k}^a (1 + jitter_{s,k}) \qquad \textbf{(10.31)}$$

The jitter for the sustain part of the fundamental can be seen in figure 10.8 in the frequency domain (top) and the time domain (bottom). The dotted line is the jitter after modification.

The resulting time varying frequency is shown in figure 10.9, with the original piano and the clean trumpet frequencies, which have been offset to facilitate reading of the plot.

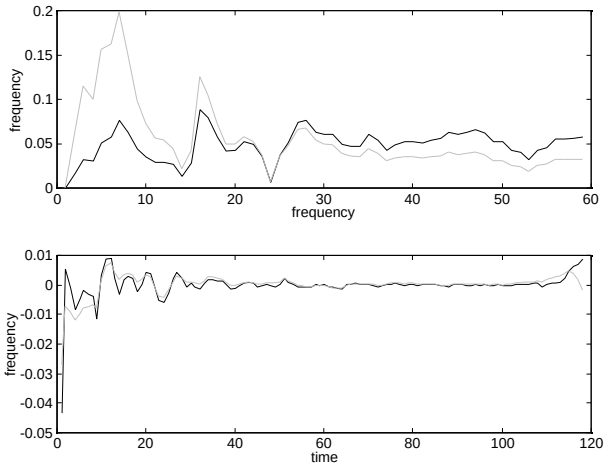


Figure 10.8. Original and modified (dotted) fundamental jitter of the piano. Frequency response (top) and time domain (bottom).

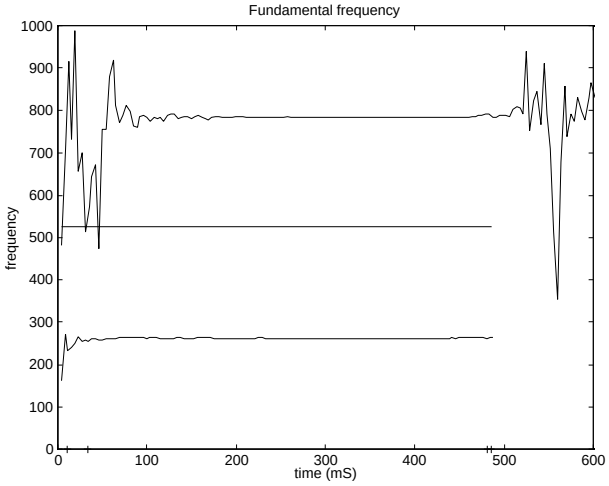


Figure 10.9. Original piano fundamental frequency (top), clean trumpet frequency (middle) and modified piano frequency (bottom). The frequencies have been offset to facilitate reading.

The jitter of the trumpet is more low-pass and it has a slightly higher standard deviation. This corresponds well with the piano standard deviation 0.1% and the trumpet standard deviation 0.4%. The filter coefficient is -0.27 for the piano and -0.72 for the trumpet. The important visual difference in figure 10.9 between the original and modified piano frequencies is due to the greater standard deviation for the piano in the attack and release

(about 4 times greater). The plus signs at the x-axis denote the split points in the trumpet sound.

Shimmer and jitter are modified for the attack, sustain and release for all partials. The correlation is not modified in this chapter.

10.5.5. Verification

The modification of the noise parameters concludes the modification of the additive parameters of a piano sound[a] into a trumpet sound[b]. It is now supposed to have the same HLA parameters as the trumpet. To compare the HLA parameters, the HLA set of the modified additive parameters of the piano sound is calculated.

The analyzed HLA parameters of the additive parameters of the modified piano are shown in figure 10.10 and the original trumpet attributes are shown in figure 10.11.

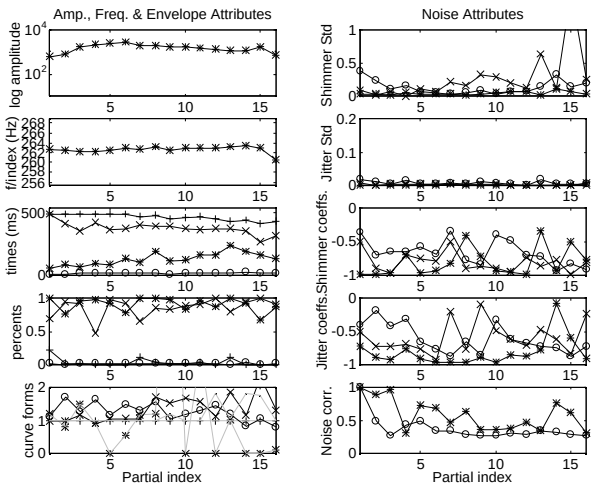


Figure 10.10. Modified piano High Level Attributes

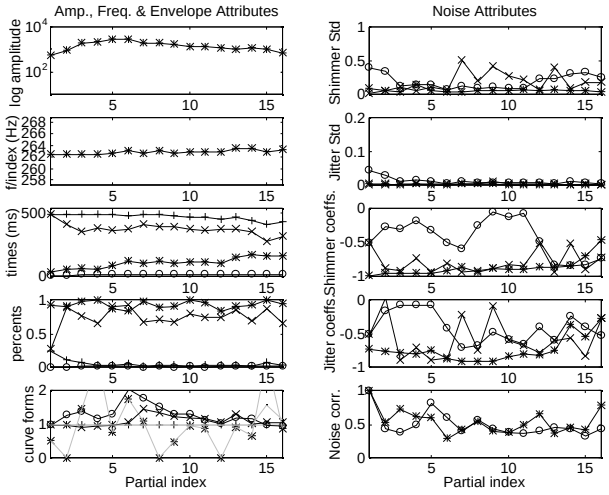


Figure 10.11. Original trumpet High Level Attributes.

The spectral envelope, frequency and envelope parameters match fairly well, whereas the noise parameters do differ somewhat. The spectral envelope and frequency values are changed by the noise components, and the other parameters are offset by different envelope time values. A slight difference in envelope times may give rise to a more important difference in the percent and in the curve form, which then changes the noise values altogether. The important envelope parameters match well, and the resynthesis of the sounds is of good quality, as seen in section 10.6 Therefore the conclusion of the additive modification is that it clearly changes the perceptive quality of sound[a] into that of sound[b].

10.6. Resynthesis

The additive parameters modified in section 10.3 can now be visualized and used for synthesis of sounds. The additive parameters for the piano before and after modifications can be seen in figure 10.12. The additive parameters in the middle clearly have the shape of the trumpet parameters to the right. The irregularities and noise on the partials is of course different, but this doesn't change the identity of the resynthesized sound.

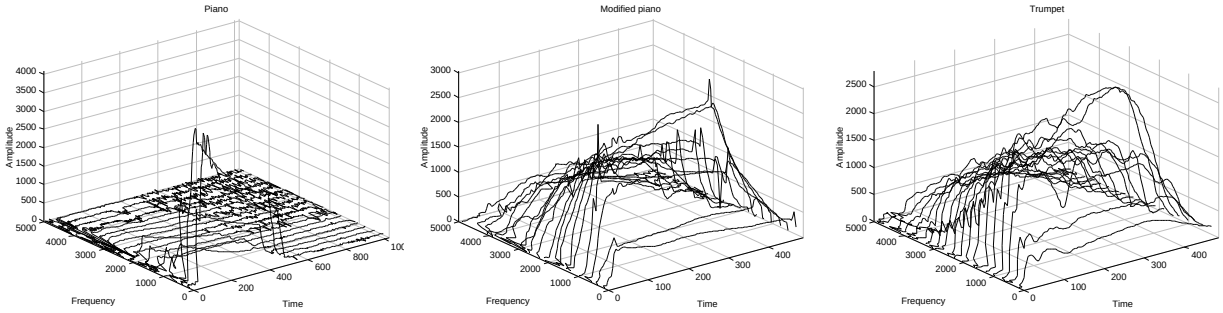


Figure 10.12. Additive parameters for the piano (left), the modified piano(middle) and the trumpet (right).

The resynthesis of many modifications of different sounds permits the conclusion that this modification is indeed very efficient, in fact so much so the original instrument in all cases is impossible to recognize. The new instrument identity seems obvious and the sound quality is generally very good. The sounds sometimes have a problem of loose harmonics, that is, harmonics that stick out of the sound. Although it is difficult to state that the sounds must have been created by the instrument it was modified into, it still sounds very close to that.

The modification of one sound into the same sound doesn't change the timbre perceptively. This shows the stability of the timbre modification method.

10.7. Conclusions

This work presents a stable method for the timbre modification with high sound quality and very good timbre identity restitution. This method has numerous applications, ranging from the creation of hybrid sounds, the control of synthetic sounds by acoustic instruments, as proposed in [Møller 1997] to the creation of sounds in inaccessible playing ranges. The method is equally fitted for the transformation of one sound into another sound, or into the same sound with another fundamental frequency, or with another intensity. The quality of the resynthesis is generally as good as the quality of the original parameters.

Modifications of the expression parameters pitch, intensity and duration of a sound is also presented, and some general indications of the modification of timbre necessary to retain a realistic sound when these parameters are changed is given.

Chapter Eleven

11. Verification of the Timbre Models

Although the timbre parameters extracted in the previous chapters seem to assist in the understanding of timbre, no formal verification of their validity has been made. The goal here to show the ability of these parameters to classify correctly a great number of sounds.

The success of the classification confirms the validity of the timbre models, and suggests the use of these models for the automatic identification of musical instruments. Furthermore, analysis of the classification ability of the different attributes helps in understanding the importance of these attributes in the perception of musical instruments.

Two different approaches are used here, principal component analysis (PCA), and classification. The PCA is helpful in understanding the significance of the different timbre attributes and a large number of sounds in the full playing range of five different instruments are classified with no errors, thus proving the validity of the timbre attributes.

11.1. Introduction

The different models of the timbre attributes are the HLA, MDA and IDA models. The HLA model presented in Chapter 6 has individual envelopes, noise, amplitude and frequency for each partial. The MDA model presented in Chapter 8 has a simple model of every MDA parameter as a function of the partial index, and the IDA model presented in Chapter 9 additionally models the evolution of all parameters from low notes to high notes.

It is here attempted to verify the validity of the different timbre models presented in this work. One way of verification is the listening tests performed in the next chapter. In this chapter, several methods are used to see if the timbre attributes classify the sounds into the instrument families correctly.

Classification using timbre attributes is a difficult task. The principal component analysis of the timbre attributes can be compared to the perceptual scaling of musical timbres [Grey 1977]. See also the dissimilarity section in Chapter 2. [Scheirer *et al.* 1997] presents a robust speech/music discriminator using brightness and other parameters. [Dubnov *et al.* 1997] shows the importance of phase coupling in the classification of musical instruments. In general terms, musical instrument classification can be compared to the task of speaker identification [Doddington 1985].

Two methods of analyzing timbre attributes are used in this work; the principal component analysis (PCA) and classification using the log likelihood [Frieden 1983]. [Skovenborg 1997] used similar methods on the time-varying amplitudes of the harmonic overtones of a small musical data set.

Other interesting methods, which are not investigated in this work, include the classification in binary trees using maximum entropy, which has proven successful in speech recognition research [Bahl *et al.* 1989], [Bahl *et al.* 1991], [Jensen 1993]. Analysis of the maximum entropy decisions used to create the binary trees could potentially give important information about the timbre attributes.

This chapter starts with a presentation of the data used in the classification in section 11.2, followed by an overview of the timbre attributes in section 11.3. Section 11.4 presents the classification using the amplitudes at the nyquist frequency of an ideal spectral envelope. The Principal Component Analysis (PCA) of the timbre attributes is presented in section 11.5. The classification of the timbre data using the log likelihood is performed in section 11.6. Finally a conclusion is offered.

11.2. Sounds

The data used in this classification are the timbre attributes of a number of sounds. These sounds come from five instruments: the piano, the violin, the clarinet, the flute and the soprano voice. The sounds used in this chapter are the same sounds that are used in the listening tests in Chapter 12.

Each instrument contributes with 30 sounds; there are thus 150 sounds in all. All of the sounds from each instrument are distributed in the normal playing range, as can be seen in figure 11.1. The sounds are played in detaché, staccato or tenuto, intensity mezzo forte. The distribution over the full playing range makes the classification of the sounds more difficult, since two sounds in the upper playing range from two different instruments are often more similar than two sounds from the same instrument with different pitch.

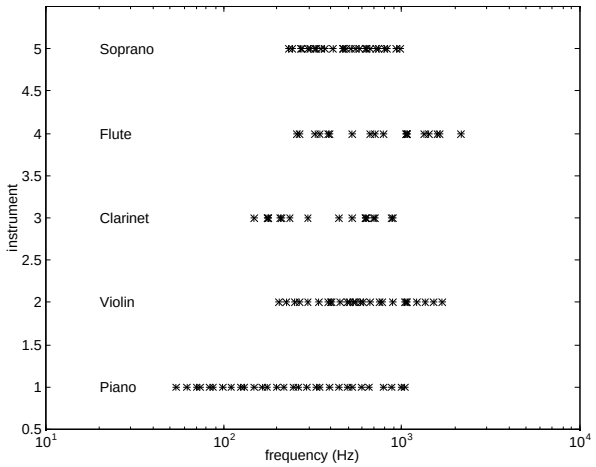


Figure 11.1. Frequency range of instruments.

11.3. Timbre Attributes

The classification is made on a subset of the MDA parameters. No automatic method for the extraction of the minimal subset necessary and sufficient for the classification of the sounds in musical instrument classes has been found. Instead a combination of the analysis of the parameters and the PCA and classification results are used to add or remove parameters, until no more parameters can be removed without degrading the classification.

A short overview of all timbre models is given here, although only the results of the MDA parameter analysis are given.

11.3.1. HLA

The HLA model is presented in Chapter 6. There are 30 HLA attributes for each partial. These are 5 envelope times, 4 envelope percents, 5 envelope curve forms, 3 noise std and 3 noise filter coefficients for shimmer and jitter, the shimmer and jitter correlation, the mean

frequency and maximum amplitude for each partial. The HLA values are not used in this chapter.

11.3.2. MDA

The MDA model is presented in Chapter 8. The MDA generally has the same parameters as the HLA, but instead of having one separate value for each partial, the MDA has a fundamental value and an exponential value. The frequency is modeled by the fundamental and the inharmonicity index, and the spectral envelope is modeled by the tristimulus 1 & 2, odd, brightness, irregularity and maximum amplitude.

In this work, only the fundamental values of each attribute are used.

Some of the attributes are related to the length or the amplitude of the sounds; these are the start length, the sustain length, the total length, and the maximum amplitude, and they are removed, because not all sounds are performed with the same duration or loudness. The fundamental frequency is also removed.

Since many of the attributes seem correlated, they are also removed. The attack and release noise values are assumed to be correlated with the sustain noise, and therefore removed.

The start segment, release segment and end segment curve forms are judged insignificant and therefore removed. The attack curve form is judged important, since the attack is one of the perceptually most important segments.

The start, attack and end percents are also removed. The start of release percents are judged important in distinguishing between sustained sounds and decaying sounds, such as the piano.

The frequency brightness is used, expressed in Hz, but all other spectral envelope attributes are expressed in partial number.

The remaining MDA attributes are used in the rest of this chapter, unless otherwise noted.

There are 16 attributes, which can further be divided into several timbre classes, spectral envelope attributes, envelope attributes, shimmer attributes and jitter attributes, as can be seen in figure 11.2.

· Spectral Envelope Attributes · Tristimulus 1 · Tristimulus 2 · Odd · Brightness · Irregularity · Envelope Attributes · Attack Time · Release Time · Start of Release Percents · Attack Curve Form	· Shimmer Attributes · Sustain Shimmer std · Sustain Shimmer filter coefficient · Shimmer Correlation · Jitter Attributes · Sustain Jitter std · Sustain Jitter filter coefficient · Jitter Correlation · Inharmonicity

Figure 11.2. Timbre attributes used in the classification.

11.3.3. IDA

The IDA model is presented in Chapter 9. The IDA values are the same as the MDA values, except that they are averaged across one half octave in 15 steps, from 32 Hz to 4 kHz. The IDA values are not used in this chapter.

11.4. Nyquist Frequency Amplitude

The nyquist frequency is the sample rate frequency divided by two. It is the highest frequency that can be analyzed with the fourier analysis. Here it is assumed to be the limit of the hearing capability, which means that frequencies above cannot be heard by the human ear. The nyquist frequency thus has a relevance to human perception. This is a simple assumption, the validity of which can be discussed. See for instance [Oohashi *et al.* 1997] for a study of the physiological and psychological effects of high frequency components.

This section will analyze the amplitudes of the ideal amplitude at the nyquist frequency. By ideal, it is assumed that the higher partial amplitudes are defined only by the brightness of the sound, and not by any resonances, or noise. Assuming that the brightness of the sound is T_b, the ideal spectral envelope is then defined by the exponential series,

$$a_k = B^{-(k-1)} \qquad \textbf{(11.1)}$$

where B is defined as (cf. Chapter 7),

$$B = \frac{T_b}{T_b - 1} \tag{11.2}$$

The partial index at the nyquist frequency is,

$$k_{nyquist} = \frac{samplerate/2}{f_0} \tag{11.3}$$

and the amplitude of the partial at the nyquist frequency is then,

$$a_{nyquist} = B^{-(k_{nyquist}-1)} \tag{11.4}$$

The frequency brightness for the five instruments are shown in figure 11.3.

The ideal spectral envelopes, that is the values of equation (11.1), calculated with the partial index brightness of five instruments, are shown in figure 11.4 and the amplitudes of the same instrument sound at the nyquist frequency are shown in figure 11.5. The instrument order is piano, violin, clarinet, flute and soprano. All axis have been normalized between instruments.

The nyquist frequency is here 16 kHz.

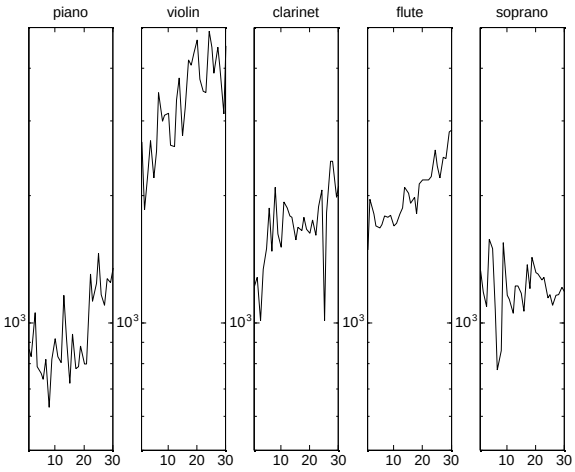


Figure 11.3. Frequency brightness for five instruments, piano, violin, clarinet, flute and soprano. x axis is sound index and y axis is frequency.

It is clear that the violin has higher amplitude at the nyquist frequency than the other instruments. The value of $a_{nyquist}$ is then capable of classifying the violin with few errors. The clarinet and the flute have the values at about 10^{-4}.The piano has the lowest values of $a_{nyquist}$ for most notes and the soprano has values in between the clarinet/flute and the piano values. This makes up four distinct groups, where most of the instruments can be positioned without errors. If the sounds are separated by the limits 5.8e-04, 1.7e-06 and 1.3e-08, the classification finds 28 violins with no errors, 57 clarinet/flute with 6 errors, 15 soprano with 7 errors and 25 piano with 12 errors. Total correct classification 125/150 with 25 errors. Remember though that this is the classification in four groups, whereas the classification in section 11.6 is done in five groups.

The values of $a_{nyquist}$ therefore are interesting in the classification of musical instruments. The values of $a_{nyquist}$ seems to give better classification than the frequency brightness in figure 11.3, which have more overlap between instruments.

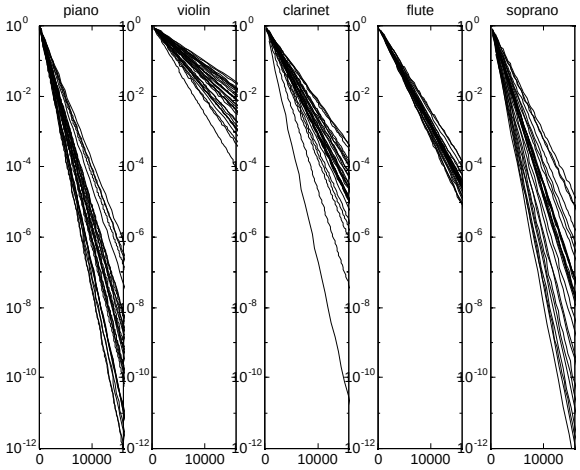


Figure 11.4. Ideal spectral envelope plotted up to nyquist for five instruments, piano, violin, clarinet, flute and soprano. x axis is sound index and y axis is amplitude.

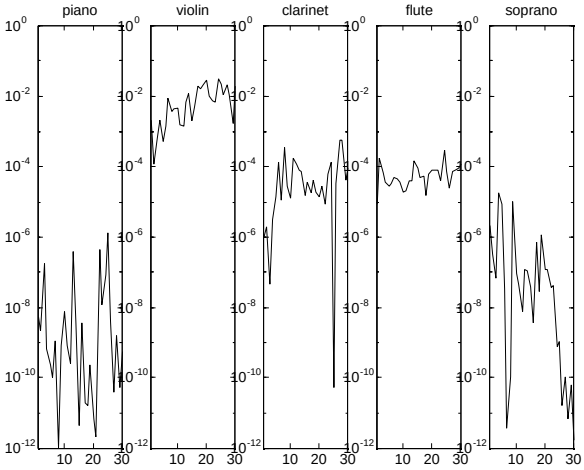


Figure 11.5. Amplitudes at nyquist for five instruments, piano, violin, clarinet, flute and soprano. x axis is sound index and y axis is amplitude.

In conclusion, the fundamental combined with brightness can give a fair classification.

11.5. Principal Component Analysis

In the Principal Component Analysis [Frieden 1983], the covariance-matrix of the timbre attributes is created, and the eigenvalues are calculated. When the eigenvalues are sorted decreasing, the eigenvectors corresponding to the *L* largest eigenvalues are enough to classify the data. Related techniques have been used for many years in the classification of musical instruments from perceptive input [Grey 1977], [Krumhansl 1989], [Iverson *et al.* 1993], [Krimphoff *et al.* 1994], [McAdams *et al.* 1995]. PCA has been used in [Sandell *et al.* 1995] for the data reduction of additive parameters. [Hourdin *et al.* 1997] uses a related multidimensional scaling for the same purpose.

Often, the 3 most prominent dimensions are used and plotted in a three-dimensional space. The separation into classes can then be verified visually.

If X is the NxM matrix of N timbre attributes, M sounds, first

$$C_x = \mathrm{cov}(X) \qquad \textbf{(11.5)}$$

is calculated and diagonalized,

The diagonal elements are sorted and assigned to λ. The value of λ indicates how much of the energy of the original timbre attributes is connected to the corresponding dimension. Therefore, if most of the energy of X is contained in the first L elements of the eigenvalues λ, then X can be transformed into the subspace spanned by the L first eigenvectors E_L.

The transformation to the eigen subspace is done by multiplying the eigenvector E_L with the original timbre attributes,

$$X = E_L \; X \qquad \textbf{(11.6)}$$

The squared error is then the sum of the L+1 to N elements in λ.

The PCA is performed on the fundamental value of the MDA attributes, described in paragraph 11.3.2, with the addition of the release curve form, and the envelope percents. The sorted eigenvalues are shown in figure 11.6. Figure 11.6 suggests that more than three dimensions are necessary to classify the data. Nevertheless, the first three dimensions are calculated and plotted. The analysis of this plot will give an idea of the distribution of the sounds in the multi-dimensional timbre space.

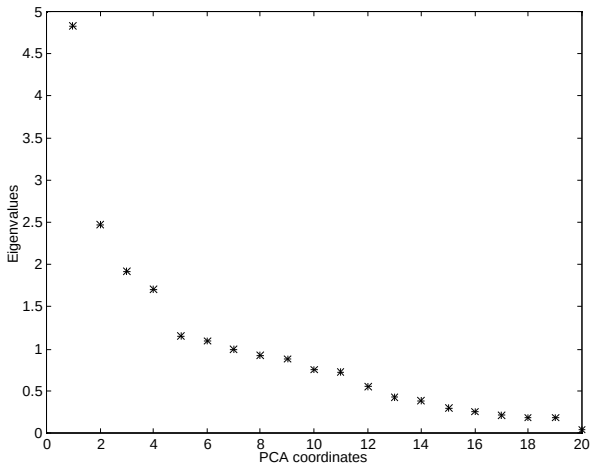


Figure 11.6. Sorted eigenvalues for the MDA timbre attributes.

The three-dimensional plot of the first 3 dimensions of the eigenvectors is shown in figure 11.7. The piano is ‘x’, the violin is ‘+’, the clarinet is ‘o’, the flute is ‘.’ and the soprano is ‘x’. Separation into the different instrument classes is quite good, although the flute doesn’t seem to be in its own group.

To facilitate reading, the data is also plotted on the three visible planes. The three-dimensional data is the one in the middle (top). The other three groups are the data plotted on the planes.

The soprano and the violin sounds are well grouped in two separate clusters. The flute also seems well separated, but the piano and the clarinet sounds are not well isolated in these three dimensions. More dimensions are necessary to separate all five instruments.

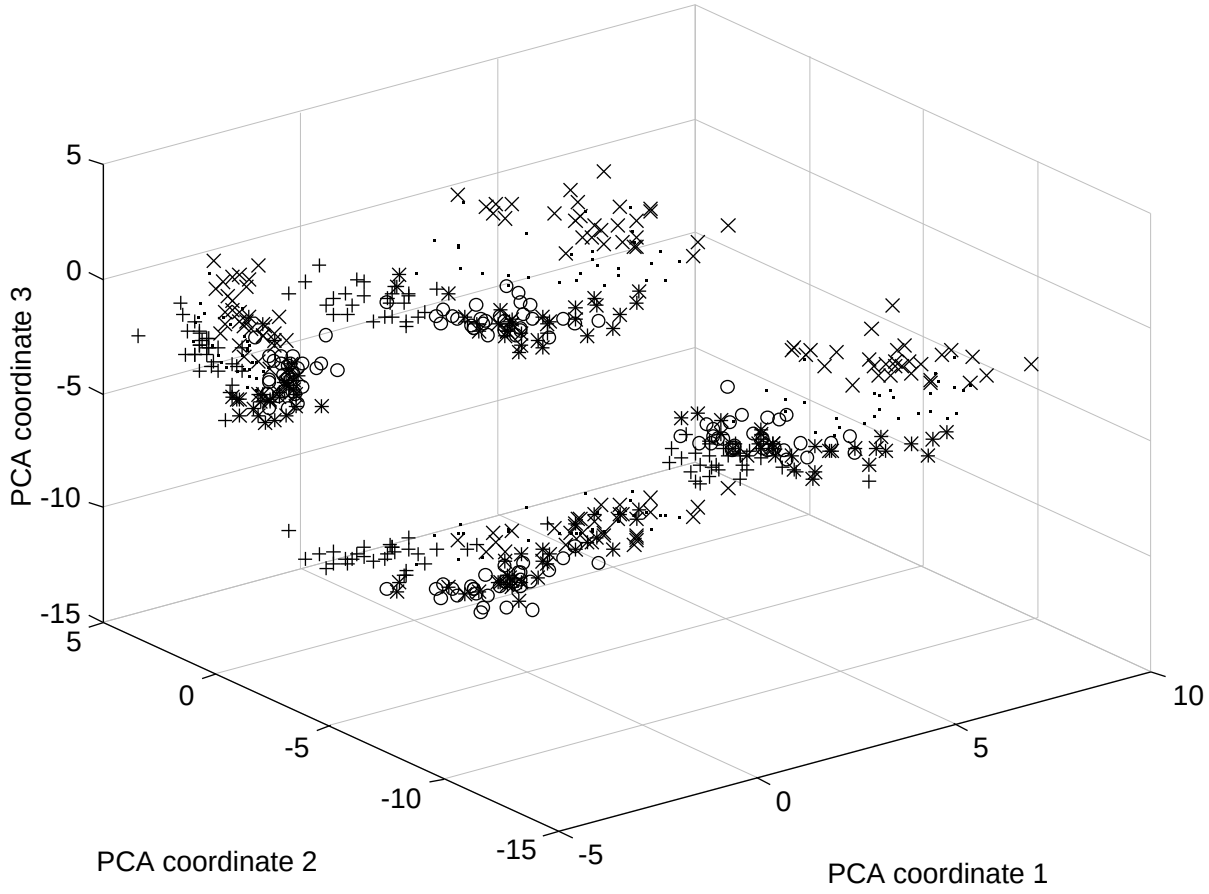


Figure 11.7. The 5 instruments in the 3 first PCA dimensions. The piano is '*', the violin is '+', the clarinet is 'o', the flute is '.' and the soprano is 'x'. Observe that the data is also plotted on the x, y and z plane. The upper cluster in the middle is the 3D plot.

It is interesting to see what attributes are prominent in the PCA coordinates. The 3 first eigenvectors are shown in figure 11.8. The first dimension is indicated with '*', the second dimension is plotted with '+' and the third dimension is plotted with 'o'.

The first dimension has prominent values (above 0.3) for sustain shimmer & jitter filter coefficients, shimmer & jitter correlation, Tristimulus 1 and Odd. This dimension seems to be principally related to the noise and the spectral envelope.

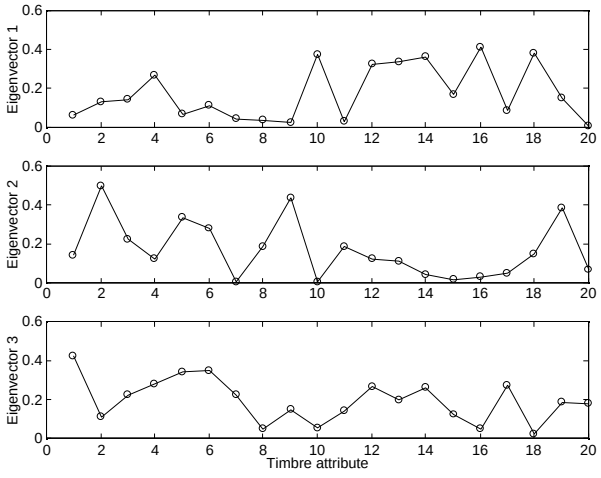


Figure 11.8. Eigenvalues for the first 3 PCA dimensions.

The second dimension has prominent eigenvector values for release time, start of release percent, sustain shimmer std and brightness. This dimension seems to be principally related to the decay envelope and brightness. The third dimension has prominent values for attack time, start of release percent and end of release percent, and is mostly related to the envelope.

11.6. Classification

In the classification, the timbre attribute data is first classified in the different instrument classes. To avoid using the same data for the classification and the test, the classification is performed using the leave one out (LOO) method, on all data except the sound being tested.

The log likelihood function for normal-distributed data is used as a classifier,

$$d_{i,k} = \frac{1}{2}\log(|C_i|) + \frac{1}{2}(X_k - m_i)^T C_i^{-1}(X_k - m_i) \qquad \textbf{(11.7)}$$

where C_i is the covariance matrix of the class i, X_k is the attributes of sound k and m_i is the mean of the class i. This model is anisotropic, which means that the shape of each class is an ellipsoid.

To ensure invertability of C_i an isotropic noise term ε is added to the diagonal elements of C_i. The value of ε is chosen in order to optimize classification.

The LOO method is as follows. The classes are created for all data except the data of the sound k that is being classified. The distance is then calculated for the five instrument classes, and the sound is placed in the class to which it has the smallest distance. This is repeated for all sounds.

If all 16 timbre attributes, as defined in paragraph 11.3.2 are used, then there is no error in the classifications. It can therefore be concluded that these 16 timbre dimensions are enough to separate the 5 instrument sounds.

Next, to verify the influence of the individual timbre attributes, the timbre attributes are divided into classes, envelope, shimmer, jitter and spectral envelope. The clustering is done for each timbre attribute class, and the number of errors are

Attribute class:	: total errors, (piano, violin, clarinet, flute & soprano)
Envelope (4 attributes)	: 55 errors, (10, 9, 8, 17 & 11)
Shimmer(3 attributes)	: 66 errors, (11, 7, 12, 12 & 24)
Jitter (3 attributes)	: 68 errors, (20, 4, 27, 9 & 8)
Spectral Envelope (5 attributes)	: 24 errors, (7, 1, 5, 3 & 8)

The piano has 48 errors cumulated, the violin has 21 errors, the clarinet has 52 errors, the flute has 41 errors and the soprano has 51 errors cumulated.

If each timbre attribute is clustered individually, then the only attribute, which has fewer than 70 errors, is brightness, which has 56 errors. All other timbre attributes yield 90+ errors. Brightness combined with the fundamental frequency yields 43 errors.

11.7. Conclusions

In this chapter, an analysis of the importance of the timbre dimensions is performed. Three methods have been tested: standard PCA, classification using the log likelihood function for normal distributed data, and a simple classification using the brightness and the fundamental frequency.

The PCA revealed some important timbre dimensions, the spectral envelope, noise and envelope being the most important. Although the PCA results are not convincing, it has been of help in choosing which timbre attributes to use in the classification.

Classification using only the brightness and the fundamental frequency by calculating the amplitude at the hearing limit of each sound gave good classification results.

The verification of the timbre model has shown that a subset of the MDA parameters is enough to classify 150 sounds from 5 musical instruments without errors. It can therefore be concluded that these attributes are important to the timbre model, since the classification was made on the same criteria as human classification of musical instruments is made. Indeed, it would seem that this classification is better than the human classification, since several subjects in the listening tests performed on the same sounds had difficulty recognizing some of the instruments, even from the original sounds.

It can tentatively be said about the attributes that were used for the classification that they can be divided into several groups: attack envelope, release envelope, noise quality, amplitude irregularity and spectral envelope. None of the subgroups can classify the sounds well, although the spectral envelope, and notably brightness, is a good classifier.

The order of importance of the timbre attribute classes in order of ability to classify might be spectral envelope, envelope and noise.

Chapter Twelve

12. Listening Tests

This chapter presents the results of listening tests performed to objectively evaluate the quality of the sounds resynthesized from the different timbre models. Five models are evaluated: the original sounds, the analysis/synthesis sounds, the HLA model sounds, the MDA model sounds, and the IDA model sounds. The results of the listening tests show that the analysis/synthesis and the HLA models generally score above annoying degradation, whereas the MDA and IDA sound quality is unacceptable. Analysis of the scores can help in improving the timbre models, or the estimation of the timbre model parameters

12.1. Introduction

In this chapter, the quality of the different timbre models is measured by subjective quality as judged by a number of listeners, called subjects.

Not many objective listening tests have been performed in the music community to evaluate synthesis methods. [Strong *et al.* 1966] evaluated the spectral/time envelope

model with listening tests. [Grey *et al.* 1977] compared analysis/synthesis and different data-reductions, and [Sandell *et al.* 1995] evaluated the PCA-based data reduction with listening tests.

The listening tests performed here have been inspired by the listening tests performed for the evaluation of speech and music compression. The method used is called double blind triple stimulus with hidden reference [ITU-R 85/10 1994]. A practical application of this test can be found in [Nielsen 1995]. The subjects are presented with three sounds, the first always being the reference and the two next sounds are the reference and the modeled sound in random order. The subjects are then asked to rate the two sounds, called B and C, against the reference sound (the original) in a scale from 1.0 to 5.0. The scale indicates the degree of impairment. The subjects are allowed to listen to each sequence again, as many times as necessary.

The test performed here differs from normal double blind triple stimulus with hidden reference tests because the sounds under test are short, and there is no use changing between sounds while listening to them, as is usual when longer music pieces are under test. Instead, the test subjects are allowed to repeat all three sounds as many times as necessary.

12.2. Rating scales

The scales used have been borrowed from [ITU-R 85/10 1994]. The impairment of the modeled sounds is judged in five steps, although the subjects are asked to give one decimal to the score when possible.

The scale is,

Score	Impairment
5.0	Imperceptible
4.0	Perceptible, but not annoying
3.0	Slightly annoying
2.0	Annoying
1.0	Very annoying

In Danish, the language of this test, this translates to [Poulsen 1996],

Vurdering	Forringelse

5.0	Ikke hørbar
4.0	Hørbar, men ikke generende
3.0	Lidt generende
2.0	Generende
1.0	Meget generende

It is stressed that what is judged are musical sounds, and what is judged is the impairment were the sounds to occur in a normal musical situation. This means, for instance, that if the subject believes that the sound is another recording from the same playing condition on the same instrument, or instrument type, which can sound quite different, but the quality is natural, then the score is 4 or above. The scores 3, 2 and 1 are used when the identity of the sound has been altered, or when the quality of the sound is deteriorated. The score 4 or higher is also used when the impaired sound is better than the original sound.

12.3. Original Sounds

There are sounds from 5 different instruments in the test: piano, violin, clarinet, flute and soprano voice. The sounds are fairly short, typically less than one second long, and they range over the normal playing range of each instrument. There are 15 sounds for each instrument. These are the same sounds that are used in the classification in Chapter 11, although every second sound is used here.

12.4. Model Sounds

There are 5 models in this test, the original, analysis/synthesis, HLA, MDA, and IDA sounds.

Analysis/synthesis is done with the linear time/frequency analysis presented in Chapter 4, spurious frequencies are analyzed, but not resynthesized and partials are smoothed over one period. The maximum number of partials is 54, and the note of each analyzed sound is given to the analysis, to reduce the influence of the fundamental frequency estimation error, as explained in Chapter 9.

The HLA model sounds are created as described in Chapter 6. The MDA model is presented in Chapter 8 and the IDA model

in Chapter 9. The MDA and IDA model sounds are made with no error term.

12.5. Listening panel

It is preferable to have only musicians, or music people, in the listening panel, because of their improved listening ability and vocabulary. 24 people are chosen to the panel, aged between 22 and 61 years.

The mean age of all subjects is 28.3 year. There are 19 male subjects and 5 female subjects, and divided into 5 non-musicians, 11 amateur-musicians and 8 musicians or music students.

12.6. Training

In training, the subjects are supposed to familiarize themselves with the sounds, the facilities, and the impairment scale. Training was done immediately prior to the test. First the subjects were presented with a paper with instructions. The content of the instructions (in Danish) can be found in appendix B.

In the first half of the training the subjects are presented with 5 typical sounds in the different modeling schemes; original, analysis/synthesis, HLA, MDA, and IDA. The first half of the training is done sequentially, and cannot be repeated.

The second half of the training is 5 normal double blind triple stimulus with hidden reference tests, where the subject can ask the supervisor questions, and the supervisor verify that the subject has understood the test procedure.

The subjects are presented with a paper with the impairment scale before the training is performed. This scale was generally only consulted in the beginning of the tests.

12.7. Test Procedure

The tests are performed with matlab [Mathworks 1992] on a Power Macintosh 7500, with a Sennheiser HD560 ovation II headphone connected to the headphone output of the Macintosh. The subjects are not allowed to adjust the amplitude, which is set to the maximum possible.

The Macintosh is placed behind a screen, but the room is not silent, and there can be some noise in the background. Although this may influence the results, it has been judged

that the impairments of the sound are so pronounced that it will not influence the judgment dramatically.

The total number of tests for each subject is 345, which is 5 models times 5 instruments times 15 sounds, minus the training sounds.

The sounds are presented in a random order. When the subject has listened to the three sounds, A, B, and C, he is asked to give an impairment score for B, and then C. When the C impairment score is validated, the subject can select the next test samples. The input is done with the numerical keyboard, and the subject can at any time press 0 to listen to the three sounds again.

The test lasts about 2 hours, and subjects are asked to take a break after 1 hour. If necessary, the second half of the test is done at a later time. This was the case for about half of the subjects.

12.8. Subject Comments

The subjects were asked to write on a paper general comments about the sound, and about the test procedure. The comments were generally related to the test procedure, the impairment scale, or the original and resynthesized sounds.

12.8.1. The Test Procedure

The test procedure didn’t get many comments, although some people complained about not being able to recognize the original instrument, others wanted clicks in the beginning and in the end removed. One subjects wanted a scrollbar, instead of the numerical input. This would probably have improved the accuracy slightly. Many subjects said they were unable to judge on a decimal scale, and consequently only gave integer scores.

12.8.2. The Impairment Scale

The impairment scale was difficult for many subjects. It seems that many people tried to fill out the scale, giving the 5 models as heard in the first part of the training 5.0, 4.0, 3.0, 2.0 and 1.0 respectively. Others had a different scale, with a jump from the HLA model to the MDA. The MDA and IDA usually performed equally badly. One subject complained that it really was two things being judged: the quality of the resynthesized sound, and how different it was to the original sound.

12.8.3. The Sounds

The sounds got more comments. The piano sounds lacked attack in the otherwise good models, and got noisy in the bad models. A few resynthesized sounds sounded better than the original according to many subjects, lacking unwanted noise, or sounding cleaner. Violin and flute sometimes had less noise, while still sounding good. One subject complained about high frequency fluctuations (jitter). This is an indication that the high partials are badly analyzed. The violin was judged too sharp by one subject, lacking spectral quality by another subject. The flute loses brilliance, or breathing noise quality, according to several subjects. The soprano was easy to judge, according to some subjects.

The normal musical situation was understood differently by the subjects; some understood it as a concert situation, where blow or hammer noises are not heard, while others understood it as an isolated situation, where all noises are heard.

That the sounds were often too short to judge was a recurrent comment.

A general consensus seems to be that even the worse model retained the identity of the sound. This is important although not well founded. Some of the sounds were so short that some subjects had difficulty identifying the instrument, even of the original sound.

The models were also identifiable by some subjects. One subject with a particularly good identification of the IDA model (almost all scores were 1 for this model) replied when asked that he found the sounds from this model sounded very good. This was not generally the case, though. Most subjects found that the bad score models sounded bad. Comments like ‘It sounds like a modem’ were heard.

12.9. Statistical Presentation

When all subjects have performed the test, the data is collected, and the rating difference, which is the rating for the modeled sound minus the rating for the reference, is calculated. This is the degradation of the model. If the degradation is larger than zero, the subject has misjudged the modeled sound. This is not supposed to happen, since most modeled sounds have a pronounced difference. The result mean degradation is then presented as a function of model type, sound type and frequency.

12.9.1. Model Degradation

In figure 12.1, the mean degradations of the 5 models are plotted, with the 95 % confidence interval.

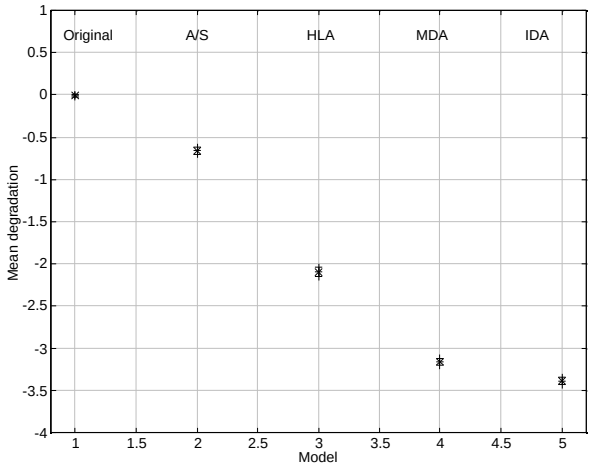


Figure 12.1. Mean degradation and 95 % confidence interval for the 5 models for all instruments and subjects.

The original sounds have zero degradation, as expected, the Analysis/Synthesis model has a perceptible, but not annoying degradation, the HLA model has a slightly annoying degradation, and the MDA and the IDA models have an annoying degradation.

Since the scale is not entirely reliable, it is perhaps more interesting to look at the relative positions of the different models.

The A/S model is positioned relatively close to the original, and the MDA and IDA models are both positioned close to the annoying degradation. The HLA model is positioned in between the A/S and MDA models, although a little lower than the exact middle. The HLA model introduces the worse degradation, which can be explained, as shown below, by the noise model of the HLA.

The next result from the listening tests is the degradation for each instrument. The analysis of the degradation can help in understanding the reasons for the low score of the timbre models.

12.9.2. Instrument Degradation

The mean degradations for the five instruments for all five models are shown in figure 12.2. The clarinet has the best score and the soprano the worst. The piano and flute also have a lower score, and the violin is next best.

The reason for the relatively worse score for the piano is probably the fast attack of the piano, which is difficult to analyze, but it could also be explained by the lower notes of the piano, and the better ability of subjects to judge low notes.

The soprano has bad scores because of vibrato on many of the sounds, which is transformed into noise in the HLA model. The flute also has some tremolo, which is not well modeled in the HLA model.

Although no screening has been performed, the score for each subject has been analyzed to understand the use of the scale of the subjects, and to see, if any subjects have a large confidence interval.

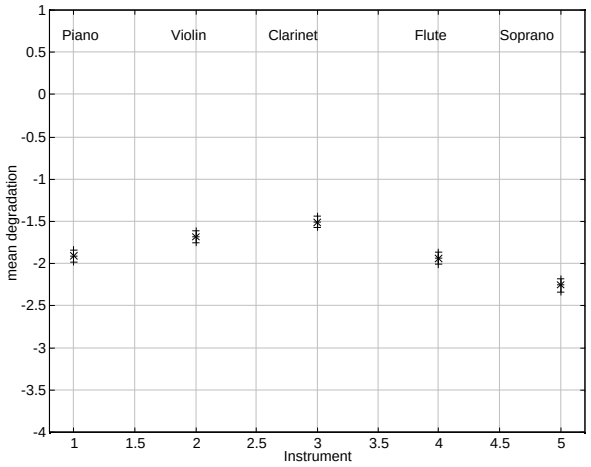


Figure 12.2. Mean degradation for the five instruments for all models and all subjects.

12.9.3. Subject Scores

The mean degradation for the subjects is shown in figure 12.3. The total degradation for each subject lies in the range between -1 and -2.5. There is a tendency for musicians to have a lower score than non-musicians, but there are exceptions.

The dispersive score for the subjects is an indication of the confusion of the scale. Another scale is probably needed for the HLA, MDA and IDA sounds, which does not try to recreate a sound, but only some of the perceptive qualities of that sound.

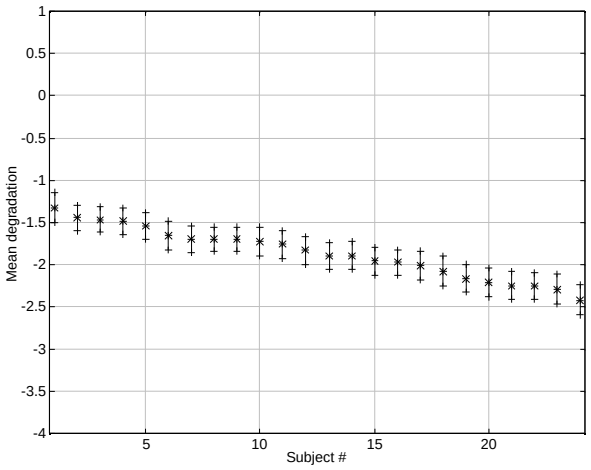


Figure 12.3. Mean degradation for the subjects for all models and all instruments.

All subjects have roughly the same confidence interval. Therefore, no screening was deemed necessary. All subjects are used in all of the degradation analysis.

The relatively worse score for the piano is now analyzed.

12.9.4. Analysis/Synthesis Instrument Degradation

In figure 12.4 the mean degradation for the five instruments for all subjects, but only for model 2 (A/S) is shown. The instrument families with fast attacks, piano and violin, perform notably worse than instruments with slower attacks.

The bad piano scores can also be explained by the relatively low pitch of many of the piano notes.

The soprano performs well in the analysis/synthesis, even though the vibrato on many of the notes is not analyzed correctly.

This can be explained by the constant frequency magnitude of the sum of the analysis filters. The constant frequency magnitude ensures that all the signal is present in the analysis, even though the partial is not positioned exactly on the frequency of the filter.

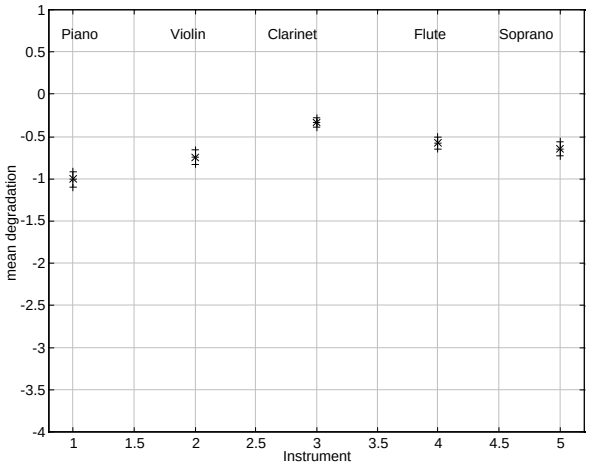


Figure 12.4. Mean degradation for the 5 instruments for al subject and model 2 (A/S).

12.9.5. Degradation as a Function of Fundamental Frequency

Figure 12.5 shows the mean degradation for the piano, model 2 (analysis/synthesis), all subjects, as a function of fundamental frequency. Degradation is clearly greater for the low fundamental frequencies. This can have several explanations: perhaps the timing resolution for the low frequencies is not good enough, and the fast attacks get blurred.

Another explanation, which remains to be verified, is that sounds with low fundamental frequency have too many partials, some of which are not analyzed.

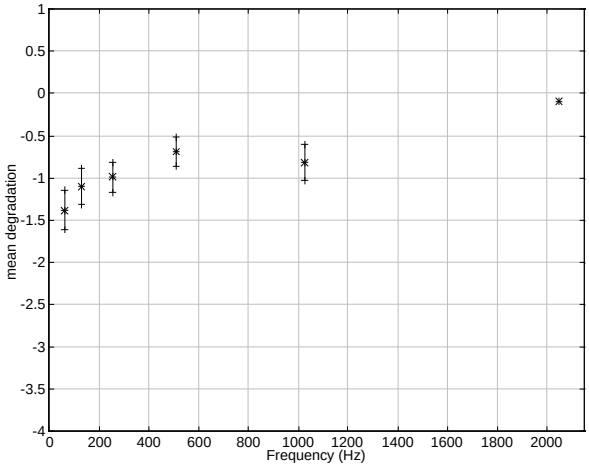


Figure 12.5. Mean degradation for the piano sound and model 2 (A/S) for all subjects, as a function of fundamental frequency.

Phase could also be influential in the bad performance of the low pitch piano notes. The phase perception is greater for low frequencies, as shown in the phase section in Chapter 2. Finally, the low piano notes have a very complex spectrum, with non-harmonic partials which are not kept in the additive analysis.

Whatever the reason, the analysis does not always perform well, getting scores approaching the annoying degradation. This should be taken into account when evaluating the HLA, MDA and IDA models, for instance by defining the mean degradation for these

models as the difference of the scores for these models and the score for model 2 (A/S). The degradation for the HLA model then approaches the perceptible, but not annoying degradation.

12.9.6. The HLA Instrument Degradations

The mean degradation for the HLA model is shown in figure 12.6. The flute and notably the soprano perform much worse than the other instruments. The reason for the flute degradation in this model might be a poor model of the flute noise, but it may also be because the flute has some tremolo effect, which is not well taken into account in the HLA model.

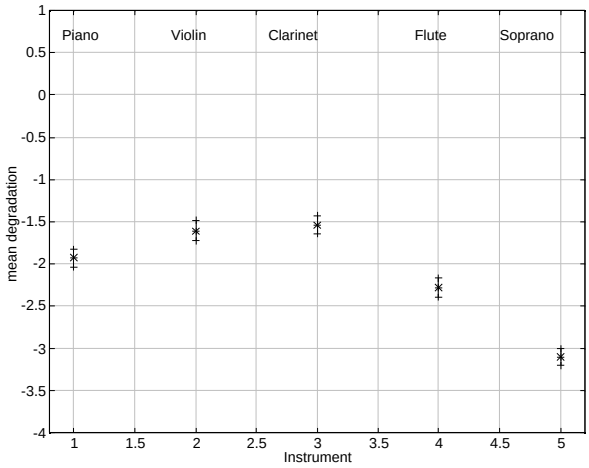


Figure 12.6. Mean degradation for the HLA model, all subjects, as a function of instrument.

The soprano definitively has a vibrato on most of the sounds.

The vibrato is not modeled in the HLA model. Small vibrato is removed, whereas large vibrato is not analyzed well in the A/S stage, and therefore translated into noise in the HLA model. The vibrato is so pronounced that it is impossible for the linear time frequency analysis to succeed in the higher partials, since they would move across several partials, when the frequency deviates. Although this is not very perceptive in the analysis/synthesis model sounds, the vibrato effect is translated into noise in the HLA model.

12.9.7. MDA Instrument Degradation

The mean degradation of the MDA model as a function of instrument is shown in figure 12.7.

The soprano still performs notably worse than the other instruments, although all of them performs worse than in the HLA model.

The degradation between the HLA and the MDA models seems to be independent of the instrument. This degradation could be attributed to the spectral envelope model, or to the correlation of the irregularities in the HLA parameters.

Furthermore, bad frequency estimation, rendering a different pitch in the MDA model could also be the cause of bad scores.

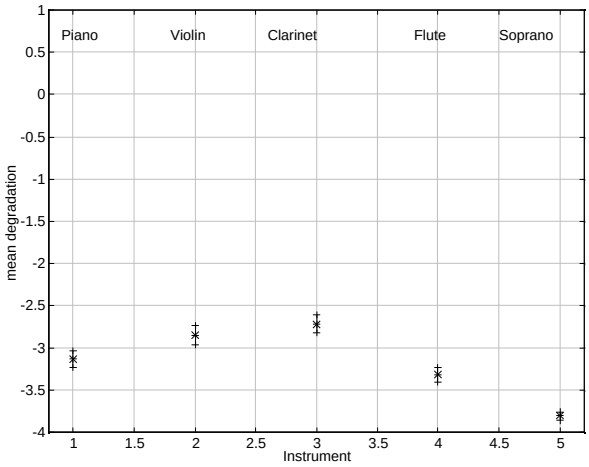


Figure 12.7. Mean degradation for the MDA model, all subjects, as a function of instrument.

12.9.8. The IDA Instrument Degradations

The IDA degradation is similar to the MDA degradation and is shown in figure 12.8. Obviously, the soprano couldn't get more degraded, much as the other instruments all fall by about the same amount.

The MDA parameter values probably need to be weighted before the mean is taken and put in the IDA model.

Again, bad pitch estimation, rendering a different pitch in the IDA model could also be the cause of bad scores.

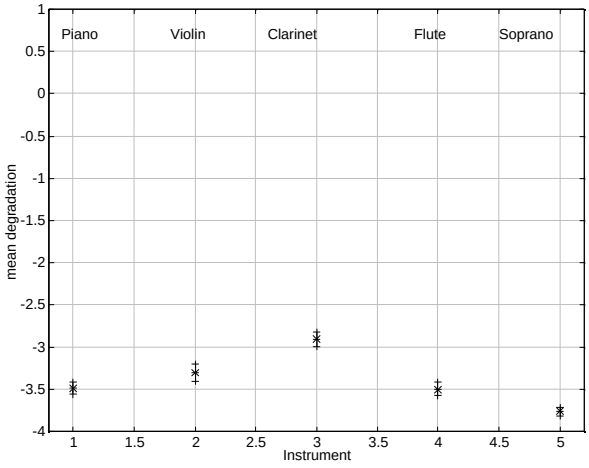


Figure 12.8. Mean degradation for the IDA model, all subjects, as a function of instrument.

Next, the relatively bad score for the soprano is analyzed.

12.9.9. Model Degradation with Soprano Removed

The mean degradation for all 5 models with the soprano removed is plotted in figure 12.9. The degradation from model 2 (A/S) to model 3 (HLA) has clearly diminished. This is due to the bad analysis of the vibrato in the soprano instrument.

The HLA, MDA and IDA scores increase by the same amount when the soprano is removed. This indicates the MDA and IDA scores are relative to the HLA score, i.e. when the HLA score changes, the MDA and IDA scores change by the same amount.

The removal of the soprano does not change the A/S score significantly.

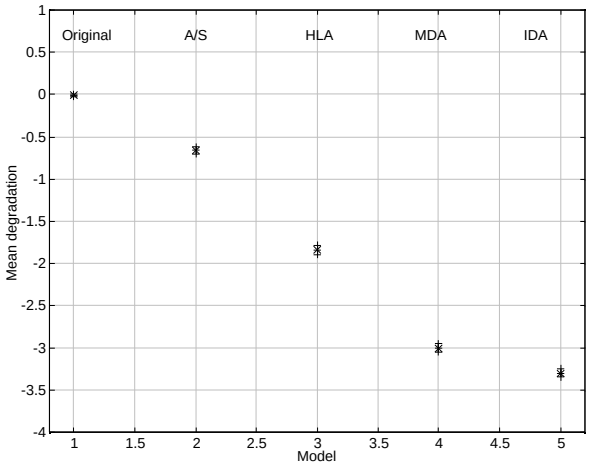


Figure 12.9. Mean degradation for the 5 models for all instruments and subjects, with the soprano removed.

12.9.10. Complete Scores

The conclusions from the preceding paragraphs can be verified by analyzing the complete scores.

The complete scores for the 5 instruments and the 5 models are shown in figure 12.10. The mean scores for all subjects are shown.

The original scores are marked with a ‘o’, the A/S scores are marked with a ‘x’, the HLA scores with a ‘*’, the MDA scores with a ‘+’, and the IDA scores are the lowest scores marked with a ‘o’. The lines between the best and the worst models indicate which instrument it is. The piano is solid, the violin is dotted, the clarinet is dashdotted, the flute is dashed, and the soprano has an empty line. The scores are sorted by the HLA scores.

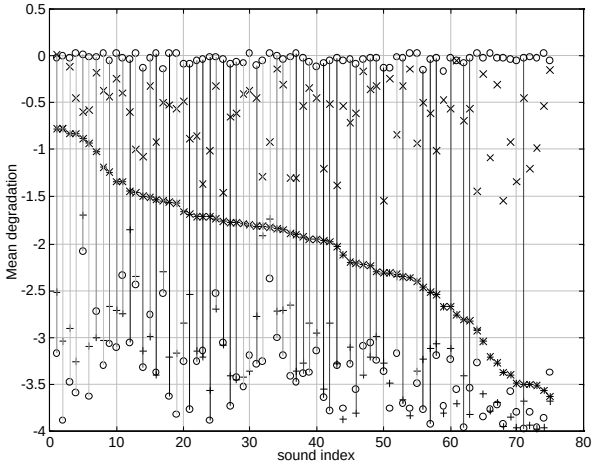


Figure 12.10. Complete scores for the 75 sounds and the 5 models.

First of all, the HLA scores varies from better than perceptible, but not annoying to almost very annoying. The violin seems to have many sounds with good HLA scores, whereas most of the worst scores are for the soprano sounds as could be expected, due to the presence of vibrato in the soprano sounds. All the scores seem to be sorted, with the original sound scores highest, followed by the A/S, the HLA, the MDA and the IDA scores. Only few of the IDA scores are better than the same scores for the MDA.

Furthermore, the instruments seem to be grouped: the violin sounds are placed in two groups, the piano also are in two groups, divided by the second violin group. Most of the soprano sounds are grouped in the lowest HLA scores. Most of the flute sounds are grouped in the second half of the figure, whereas the clarinet sounds seem more scattered in the first half of the plot.

The best HLA score sounds have a perfect A/S score, and the worst HLA score sounds have an even worse MDA and IDA score. Generally, the good HLA score sounds seems to have good A/S, MDA and IDA scores, and the bad HLA score sounds seem to have bad A/S, MDA and IDA scores.

12.10. Conclusions

The listening tests have been performed for enough subjects under good conditions. An improvement would have been to have a scrollbar, instead of the numerical input. Furthermore, the scale used in this test might not be appropriate, since there actually are two things under test: the quality of the resynthesized sound and the difference between the original and the resynthesized sounds. One modification to the degradation scale would be to have degradation 1 be ‘Unrecognizable’, and degradation 2 be ‘Very annoying’, or add a 0 degradation ‘Unrecognizable’.

The listening tests have shown that the analysis/synthesis performs above ‘Perceptible, but not annoying’, except for the low frequency piano sounds. The HLA model performs significantly better than the MDA and the IDA models and generally above slightly annoying.

The listening tests have been instrumental in understanding the reasons for the sound degradation in the timbre models. The results presented here can be used to further improve the timbre models, and the estimation of the timbre model parameters.

Chapter Thirteen

13. Conclusions

In this chapter the methods and results in the preceding chapters are summarized. The main accomplishment of this work has been the construction of a model of the timbre of isolated musical instrument sounds. Several new methods or improvements of existing methods for the estimation of the parameters of the timbre models have also been presented. The timbre models can be used to resynthesize sounds, and they are useful when analyzing timbre evolution as a function of pitch, loudness, tempo or style.

The timbre models have been evaluated by performing listening tests on the resynthesis of sounds from the parameters of the model, and by performing classification of sounds in instrument classes using the parameters of the model. Finally, timbre modification methods, which permit “playing” the models, have also been presented.

13.1. The Timbre Models

The general goal of this work was to find a model of the timbre of isolated quasi-harmonic musical sounds. This model, the High Level Attributes (HLA) model, was

presented in Chapter 6. The HLA model analyses the additive parameters and extracts pertinent, intuitive parameters.

The analysis of the additive parameters is summarized in Chapter 4. Two methods for the analysis of musical sounds were compared, the FFT-based analysis and a new method, developed by Philippe Guillemain [Guillemain *et al.* 1996], which is here called the Linear Time Frequency (LTF) analysis method. A comparison of the two methods reveals that the LTF analysis has a time resolution that is twice as good as the FFT-based analysis. The LTF analysis is therefore used in the rest of this work. The LTF analysis necessitates an estimation of the frequencies to analyze. This estimation is performed in Chapter 3. An improved fundamental frequency estimator, which estimates the fundamental of stretched harmonic sounds is presented in that chapter.

The HLA parameters consist of the spectral envelope, which is the maximum of each partial, the mean frequency of each partial, a simple envelope function and noise parameters.

The envelope model consists of five segments, start, attack, sustain, release and end for each partial. Each segment has a start and an end relative value, the start and end times and a value of the curve form (exponential/logarithmic) of the segment. The envelope model is presented in Chapter 5. A new estimation of the envelope times based on the analysis of the derivative of the envelope is presented, which performs better than a widely used percents-based method of estimating the times. The envelope model introduced in this work, which has variable split-point amplitudes, models equally well attack-decay-release (percussive) and attack-sustain-release (sustained) sounds.

Noise is modeled as the irregularity at the amplitude (shimmer) and frequency (jitter) of the partials. This noise model seems to recreate both correlated and additive noises well. The shimmer or jitter of each partial is modeled with the standard deviation and the filter coefficient of a simple filter, which have the same magnitude response as the noise. The noise model is presented in Chapter 6.

The HLA model has few intuitive parameters, and it can resynthesize an analyzed sound with high quality. The sounds are rarely identical, though, since the simple envelope model cannot recreate the amplitude variations faithfully. Nevertheless, the timbre identity of the sound is recreated flawlessly.

The HLA model still has some drawbacks. The size of the model is fixed, except for the number of partials. While this is rarely a problem, it makes comparing a high pitched flute

with five partials with a low piano sound with fifty partials difficult. The important spectral envelope also needs a model, if timbre morphing is to be performed. The morphing between two resonances in the spectral envelope should ideally move the resonance from the first position to the second. This is not the case with the spectral envelope, which also lacks simple, intuitive understanding. Furthermore, it should be possible to visualize most parameters with one parameter, which could be the fundamental value.

For these reasons, the Minimal Description Attribute (MDA) model was developed. It is presented in Chapter 8. The main feature of the MDA model is the spectral envelope model, which is presented in Chapter 7. The spectral envelope model parameters are brightness, the odd value, tristimulus and irregularity. Brightness is a measure of the mean of the spectrum; a low value indicates much amplitude in the fundamental whereas a high brightness value indicates strong high partials. Brightness is highly correlated with the perceptual quality brightness. The odd value is a measure of the amplitude of odd partials. Tristimulus is the measure of the amplitudes of three groups, the first consisting of the fundamental, the second of the first three overtones, and the third of the remaining overtones. Irregularity is a measure of the difference in amplitude between adjoining partials. These values are calculated for the spectral envelope, and this work presents a method that recreates a spectral envelope with the same spectral envelope model parameter values.

The fundamental and inharmonicity model the frequencies of the partials.

The value of the fundamental and an exponential parameter define the partial index evolutions of the other parameters, which are the envelope times, percents and curve forms and the shimmer and jitter standard deviation, filter coefficients and correlations.

The MDA model can model only quasi-harmonic sounds, because of the structure of the parameters. Some evidence that the estimation of the MDA model parameters could be improved is given in Chapter 8.

The MDA model seems well adapted for isolated musical sounds. It solves some of the problems of the HLA model, although the sound quality of the resynthesis from the MDA model is significantly lower than the resynthesis from the HLA model parameters. The spectral envelope now has an intuitive model and most parameters have a fundamental value, which can be used when visualizing the timbre attributes. However, the HLA, or MDA models can only model one sound. Most musical instruments have a pitch range, intensity range and several styles.

These can be modeled in the Instrument Definition Attributes (IDA) model. The IDA model parameters are the same as the MDA model, but they are collected for all pitches of an instrument in 15 half-octave bands. The IDA model is introduced in Chapter 9. The analysis of the evolution of the IDA model parameters is also presented in that chapter, for the fundamental frequency evolution, but also for different loudnesses, tempi and style. The IDA model facilitates the analysis of both spectral envelope, inharmonicity, envelope and noise parameters. Some of the conclusions from the analysis of the IDA parameters are that the partial index brightness decreases with the fundamental frequency, giving most of the amplitude to the fundamental for the highest fundamental frequencies. The attack time also decreases with the fundamental frequency, reaching as much as one fourth of the time of the low notes in the high notes. The intensity increase translates into an increase in both amplitude and brightness. Tempo change is seen most in the sustain and release length, as could be expected.

13.2. Timbre Modifications

Timbre modifications are presented in Chapter 10. Methods for the modifications of the important expression parameters pitch, duration and loudness are given, as well as timbre morphing methods for the different timbre models. When changing the pitch, duration or loudness of a sound, many other parameters also need to be modified, in order to retain the realism of the sound. This is explained in Chapter 10.

The modification of the better quality additive parameters is presented in detail. The modification of the additive parameters can be done with the HLA model as a template. The template can be another sound, or an interpolation between two sounds. If the interpolation is chosen, the MDA model can be used to interpolate some, or all of the parameters. All timbre attributes can be changed, be it spectral envelope, frequencies, envelope, or noise.

The sound quality of the modification of the additive parameters is of good quality. Gradual changes of most timbre attributes can be made with consistent perceptual result. This makes this work suitable for the timbre scale composition [Wessel 1979]. The better quality of the additive parameters also makes it interesting to use this timbre modification method, if the perceptual effect of different timbre attribute changes is to be analyzed.

13.3. Timbre Model Evaluation

The timbre model evaluation is done by two different approaches. The first approach involves the classification of the sounds in instrument classes. If the classification can be made with few timbre attributes, it is an indication of the pertinence of the timbre attributes.

The classification was performed using the log likelihood for normal distributed data. A subset of the timbre attributes was found, by trial and error, and by analyzing the results of a Principal Component Analysis (PCA). The PCA revealed the importance of the spectral envelope model, the attack time, the release percents and most noise parameters. Only the fundamental values of the MDA model were used in the classification.

150 sounds from five musical instruments, piano, violin, clarinet, flute and soprano, in the full playing range of each instrument, were classified with no errors. 16 parameters were used in the classification, and the order of importance of the timbre attribute classes is estimated to be spectral envelope, amplitude envelope and noise.

Another evaluation method of the timbre models was performed by asking listeners, who are called subjects, to compare the recreated sounds with the originals and judge the impairment of the resynthesis. This evaluates the quality of the resynthesis of the models, but it does not always confirm or infirm the validity of the model, since the bad quality of the resynthesis also can be attributed to problems with the estimation of the parameters of the model.

The impairment of the additive analysis is better than perceptible, but not annoying, except for the low piano notes. This problem is attributed to either bad timing resolution, or the lack of phase information in the additive model. Other causes could also be too few partials, or the lack of spurious partials, which model the transient behavior of, for instance, the piano attack.

The HLA model impairment was generally better than slightly annoying, except for the soprano. The reason for the bad score of the soprano sounds is the vibrato present in these sounds. The vibrato is so important, the sounds cannot be analyzed correctly and the vibrato is interpreted as noise, which degrades the sounds considerably in the resynthesis.

The MDA and the IDA resynthesis are comparable, but the MDA model almost always scores just above the IDA model. The impairments of these two models are in between annoying and very annoying. Some of the problems with the MDA and IDA models lie

with the noise parameters. The low strong amplitudes often inherit the noise from the upper weak partials. This can be solved by weighting the parameters before the curve fit, or by using another model of the partial index evolution of the noise parameters.

13.4. Future Directions

The problems with the timbre models presented in this work are related to the estimation of the model parameters, and the validity of the model. The validity of the model relates principally to sounds that have not been discussed in this work, such as sounds with vibrato or tremolo, speech sounds, etc.

13.4.1. Parameter Estimation

The timbre models introduced in this work are believed to be valid and pertinent. Several new methods for the estimation of the model parameters have been presented in this work. However, some problems still persist. In the additive model, the relative phase of the partials is not saved with the frequency and the amplitude. Some evidence exists that this is indeed important for the quality of the resynthesis. Furthermore, phase coupling has also been shown to be a good classification parameter [Dubnov *et al.* 1997]. The importance of phase should therefore be evaluated, and the phase should perhaps be included in all the timbre models.

The additive parameter analysis should be improved to also handle vibrato or glissando. Several methods of accomplishing this have been evaluated. The estimation of the initial frequencies using a pitch tracker has shown some promising results. This method is not ready for automatic analysis, however. The pitch track is a difficult problem, and more work is needed before this method can be put into use. Initial evaluation of the possibility of analyzing the frequency content using spectrograms has also been done. This method could be improved by the techniques found in the scale-space community in the vision research. Finally, the linear time frequency (LTF) implementation used in this work could be extended to also handle varying frequency.

The improvement of the analysis would solve some of the noise problems in the HLA model. This problem consists of vibrato or glissando being transformed into noise. Periodic noises are removed in the noise analysis, but if vibrato is important, the additive analysis does not perform well, and the HLA analysis does not get good parameters to analyze.

The spectral envelope and the frequency model seem to work well for the sounds analyzed. The envelope model also performs well, although long sounds seem to have higher shimmer values. The noise model is rather simple, and seems to be the attribute that is causing the most impairment in the resynthesis. One improvement would be to include higher order statistic models, such as skewness or kurtosis [Press *et al.* 1997].

The main problem with the Minimum Description Attribute (MDA) model is the noise parameter estimation. This problem could be solved easily, by using a different weight, or another model. Analysis of higher quality noises, such as whisper, should be performed to see if the noise model handles spectral envelope information. If not, a model similar to the spectral envelope model should be introduced for the noise standard deviations.

Another important issue in the MDA model is the parameter relationship, hereby meaning the departures from the curves found in the MDA of the parameters of the HLA model. Even if most parameters are well analyzed, and fit the MDA model, the MDA resynthesis is rarely as realistic as the HLA model resynthesis. This has to do with the parameter relationship. The MDA resynthesis would be improved if the error model of the MDA could incorporate these relationships.

The IDA model is dependent on the quality of the MDA model. One problem with the IDA model is the summation of parameters from several sounds into the same IDA frequency band. If one sound timbre attribute values are heavily off it could impair the mean of the parameters. An illustrative example would be the jitter standard deviation. If the jitter is close to zero for a few sounds, but very large for one sound, the resulting IDA value would gain too much importance, resulting in a too noisy sound. Such situations must be prevented in the IDA model parameter estimation. It impairs not only the resynthesis of the sounds, but also the analysis of the parameters.

13.4.2. Model Scope

The timbre models presented in this work can handle most sounds (the additive model), or most isolated sounds (the HLA model), or handle only quasi-harmonic sounds (the MDA and IDA models). The additive and HLA models could potentially handle noise and non-harmonic sounds.

The HLA model handles only isolated sounds, but they could be very noisy without needing additions to the HLA model. Typical expression features, such as vibrato, or tremolo, cannot be modeled with the HLA model. This would not introduce a major

change in conception, since the major problem is the lack of analysis tools. Furthermore, most expression parameters can and should be added by the performer at synthesis.

However, the vibrato effect is quite complex [Mellody *et al.* 1997] and the relations between timbre attributes need modeling, if a faithful, good quality vibrato is to be created.

The singing voice also needs an improved model, at least for the MDA and the IDA models. Related research has already proven the validity of the additive model [McAuley *et al.* 1986] or models similar to the HLA model [Marques *et al.* 1994] in the modeling of speech. Initial studies of a formant model has shown promising results, and this could be a valuable addition to the MDA and the IDA models, which would permit the modeling of formantic structures in the spectral envelope.

In a larger scope, the timbre models should handle other musical instruments, such as percussive instruments, carillons, etc. The HLA model can probably model these instruments well, if the additive parameters are correct, but the MDA and IDA models can handle only quasi-harmonic sounds. Analysis of the frequency relationship in these instruments could potentially find suitable models of the frequencies of non-harmonic sounds.

Further on, all kinds of timbre attributes of concrete sounds [Shaeffer 1966] should be incorporated in the timbre models. Industry noises and animal sounds, for instance, are considered musical sounds by many.

14. References

[Allen 1977] J. B. Allen, Short term spectral analysis, synthesis and modification by discrete fourier transform. IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-25, No. 3, June 1977.

[Ando *et al.* 1993] S. Ando, K. Yamaguchi, Statistical study of spectral parameters in musical instrument tones. J. Acoust. Soc. Am 94(1), July 1993.

[Arcos *et al.* 1997] J. L. Arcos, R. L. de Mantaras, X. Serra, SaxEx: A case-based reasoning system for generating expressive musical performances. ICMC proc. 1997.

[Arfib 1978] D. Arfib, Digital synthesis of complex spectra by means of multiplication of nonlinear distorted sine waves. J. Acoust. Soc. Am. 27(10) October, 1978.

[ASA 1960] American Standard Association, Acoustical Terminology, New York, 1960.

[Backus 1970] J. Backus, The acoustical foundation of music. John Murray Ltd. London, 1970.

[Bahl et al. 1989] L. R. Bahl, P. F. Brown, P. V. de Souza, R. L. Mercer, A tree-based statistical language model for natural language speech recognition. IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-37, No. 7, July 1989.

[Bahl et al. 1991] L. R. Bahl, S. Das, P. V. de Souza, M. Epstein, R. L. Mercer, B. Merialdo, D. Nahamo, M. A. Picheny, J. Powell, Automatic phonetic baseform determination. Proc. ICASSP, 1991.

[Barrière *et al.* 1991] J-P Barrière, Le timbre, métaphore pour la composition (collection of articles), C. Bourgois, Editor, IRCAM 1991.

[Beauchamp 1982] J. Beauchamp, Synthesis by spectral amplitude and "Brightness" matching of analyzed musical instrument tones. J. Acoust. Eng. Soc., Vol. 30, No. 6. 1982.

[Benade 1973] A. H. Benade, The physics of brasses. Scientific American, July 1973.

[Benade *et al.* 1988] A. H. Benade, S. N. Kouzoupis, The clarinet spectrum: Theory and experiment. J. Acoust. Soc. Am. 83(1), January 1988.

[Benade 1990] A. H. Benade, Fundamentals of musical acoustics. Dover publications inc. New York, 1990.

[Bernstein *et al.* 1976] A. D. Bernstein, E. D. Cooper, The piecewise-linear technique of electronic music synthesis. J. Audio Eng. Soc. Vol. 24, No. 6, July/August 1976.

[von Bismarck 1974a] G. von Bismarck, Timbre of steady sounds. Acustica, Vol. 30, 1974.

[von Bismarck 1974b] G. von Bismarck, Sharpness as an attribute of the timbre of steady sounds. Acustica, Vol. 30, 1974.

[Boashash 1992] B. Boashash, Estimating and interpretating the instantaneous frequency of a signal - Part 1: Fundamentals. Proc. of the IEEE, Vol. 80, No. 4, april 1992.

[Bode 1984] H. Bode, History of electronic sound modification. J. Acoust. Soc. Am. Vol. 32, No. 10, October 1984.

[le Brun 1979] M. le Brun, Digital waveshaping synthesis, J. Audio Eng. Soc. 27(4), April 1979.

[Buunen 1976] T. J. F. Buunen, On the perception of phase differences in acoustic signals, Doctoral Dissertion, University of Delft, The Netherlands, 1976.

[Cadoz *et al.* 1984] C. Cadoz, A. Luciani, J. Florence, Responsive input devices and sound synthesis by simulation of instrumental mechanisms: The Cordis system, Computer Music Journal 8(3), 1984.

[Cadoz *et al.* 1990] C. Cadoz, L. Lisowsky, J-L. Florens, A modular feedback keyboard design. Computer Music Journal, Vol. 14, No. 2, summer 1990.

[Charbonneau 1981] G. R. Charbonneau, Timbre and the perceptual effects of three types of data reduction. Computer Music Journal, Vol. 5, No. 2, 1981.

[Chen *et al.* 1996] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit. Dept. of Statistics Technical Report, Stanford University, February 1996.

[Cheung *et al.* 1996] N-M. Cheung, A. B. Horner, Group synthesis with genetic algorithms. J. Audio Eng. Soc. Vol. 44, No. 3, March 1996.

[Chowning 1973] J. M. Chowning, The Synthesis of complex audio spectra by means of frequency modulation. J. Acoust. Soc. Am. 21(7) September, 1973.

[Conklin 1997] H. A. Conklin, Piano strings and ‘phantom’ partials, J. Acoust. Soc. Am., Vol. 102, No. 1, 1997.

[Dennis *et al.* 1990] J.E. Dennis & R.B. Schabel, Numerical methods for unconstrained optimization and nonlinear equations, pp. 218-229. Prentice-Hall, 1990.

[Depalle *et al.* 1993] P. Depalle, G. Garcia, X. Rodet, Tracking of partials for additive sound synthesis using hidden markov models. Proc. of the IEEE, 1993.

[Ding *et al.* 1997] Y. Ding, X. Qian, Processing of musical tones using a combined quadratic polynomial-phase sinusoid and residual (QUASAR) signal model, J. Audio Eng. Soc. Vol. 45, No. 7/8, July/August 1997.

[Doddington 1985] G. R. Doddington, Speaker recognition – identifying people by their voices. Proc. of the IEEE, Vol. 73, No. 11, November 1985.

[Dorkan *et al.* 1994] E. Dorkan, S. H. Nawab, Improved musical pitch tracking using principal decomposition analysis, Proc. IEEE, 1994.

[Doval *et al.* 1991] B. Doval. X. Rodet, Estimation of fundamental frequency of musical sound signals, Proc. ICASSP, Vol. 5, May 1991.

[Dubnov *et al.* 1996] S. Dubnov, N. Tishby, D. Cohen, Investigation of frequency jitter effect on higher order moments of musical sounds with application to synthesis and classification. Proc of the Int. Comp. Music Conf. 1996.

[Dubnov *et al.* 1997] S. Dubnov, X. Rodet, Statistical modeling of sound aperiodicity. Proc of the Int. Comp. Music Conf. 1997.

[Eaglestone *et al.* 1990] B. Eaglestone, S. Oates, Analytic tools for group additive synthesis. Proc. of the ICMC, 1990.

[Fant *et al.* 1985] G. Fant, J. LiljenCrants, Q. Lin, A four-parameter model of glottal flow. Speech Transmission Laboratory Quarterly Progress Report 4/85, KTH, pp1-3. 1985.

[Fitz *et al.* 1995] K. Fitz, L. Haken, Bandwidth enhanced modeling in Lemur. Proc. of the ICMC, 1995.

[Fitz *et al.* 1996] K. Fitz, L. Haken, Sinusoidal modeling and manipulation using Lemur. Computer Music Journal, winter 1996.

[Fletcher 1964] H. Fletcher, Normal vibrating modes of a stiff piano string, J. Acoust. Soc. Am., Vol. 36, No. 1, 1964.

[Fletcher *et al.* 1991] N. H. Fletcher, T. D. Rossing, The physics of musical instruments, Springer-Verlag, 1991.

[Freed *et al.* 1997] A. Freed, T. Jehan, Database of challenging musical sounds for evaluation and refinement of pitch estimators, Proc. ICMC, Thessaloniki, Greece, 1997. See also http://cnmat.cnmat.berkeley.edu/ICMC97/papers-html/Pitch.html

[Freedman 1967] M. D. Freedman, Analysis of musical instrument tones. J. Acoust. Soc. Am. Vol. 41, No. 4, 1967.

[Friberg 1991] A. Friberg, Generative rules for music performance: A formal description of a rule system. Computer Music Journal, Vol. 15, No. 2, summer 1991.

[Friberg *et al.* 1991] A. Friberg, L. Fryden, L-G. Bodin, J. Sundberg, Performance rules for computer-controlled contemporary keyboard music. Computer Music Journal, Vol. 15, No. 2, summer 1991.

[Frieden 1983] B. R. Frieden, Probability, Statistical Optics, and Data Testing: A problem solving approach. Springer-Verlag, 1983.

[Gersho 1994] A. Gersho, Advanced speech and audio compression. Proc. of the IEEE, Vol. 82. No. 6, June 1994.

[Gonçalvès *et al.* 1998] P. Gonçalvès, E. Payot, Adaptive Diffusion Equation for Time-Frequency Representations. Proc. of the Eighth IEEE Digital Signal Processing Workshop, Bryce Canyon Nat. Park, Utah, USA, August 1998.

[Gordon 1987] J. W. Gordon, The perceptual attack time of musical tones. J. Acoust. Soc. Am. 82(2), July 1987.

[Grey 1977] J. M. Grey, Multidimensional perceptual scaling of musical timbres. J. Acoust. Soc. Am., Vol. 61, No. 5, May 1977.

[Grey *et al.* 1977] J. M. Grey, J. A. Moorer, Perceptual evaluation of synthesized musical instrument tones, J. Acoust. Soc. Am., Vol. 62, No. 2, August 1977.

[Grey *et al.* 1978] J. M. Grey, J. W. Gordon, Perceptual effects of spectral modification on musical timbres. J. Acoust. Soc. Am. Vol. 63, No. 5, May 1978.

[Gribonval *et al.* 1996] R. Gribonval, P. Depalle, X. Rodet, E. Bacry, S. Mallat, Sound signal decomposition using a high resolution matching pursuit. Proc. ICMC, 1996.

[Guillemain 1994] P. Guillemain, Analyse et modélisation de signaux sonores par des représentations temps-frequence linéaires. Ph.D. dissertion. Université d'aix-marseille II, 1994.

[Guillemain *et al.* 1996] P. Guillemain, R. Kronland-Martinet, Characterization of acoustic signals through continuous linear time-frequency representations. Proc. of the IEEE, Vol. 84, No 4, April, 1996.

[Haar Romeny *et al.* 1994] B. M ter Haar Romeny, W. J. Niessen, J. Wilting, L. M. J. Florack, Differential structure of images: accuracy of representation. Proc. of the IEEE, 1994.

[Hall *et al.* 1987] D. E. Hall, P. Clark, Piano string excitation IV: The question of missing modes. J. Acoust. Soc. Am. 82(6) 1987.

[Handel 1989] S. Handel, Listening, an introduction to the perception of auditory events. MIT press, London, 1989.

[Harris 1978] F. J. Harris, On the use of windows for harmonic analysis with the discrete fourier transform. Proc IEEE, Vol. 66, No. 1, January 1978.

[Hasbroucq *et al* 1986] T. Hasbroucq, Y. Guiard, Response determination in tactile motor tasks: Body vs. device-centered cues. Cahiers de Psychologie Cognitive, 6(4), pp. 367-377, 1986.

[Hasbroucq *et al* 1989] T. Hasbroucq, Y. Guiard, S. Kornblum, The additivity of stimulus-response compatibility with the effects of sensory and motor factors in a tactile choice reaction time task. Acta Psychologica, 72, pp. 139-144, 1989.

[Hoopen *et al.* 1982] G. ten Hoopen, S. Akerboom, E. Raaymakers, Vibrotactual choice reaction time, tactile receptor systems and ideomotor compatibility. Acta Psychologica, 50, pp. 143-157, 1982.

[Horner *et al.* 1993] A. Horner, J. Beauchamp, L. Haken, Machine tongues XVI: Genetic algorithms and their application to FM matching synthesis. Computer Music Journal, winter 1993.

[Horner *et al.* 1995] A. Horner, J. Beauchamp, Synthesis of trumpet tones using a wavetable and a dynamic filter. J. Audio Eng. Soc. Vol. 43, No. 10, October 1995.

[Horner *et al.* 1996] A. Horner, J. Beauchamp, Piecewise-linear approximation of additive synthesis envelopes: A comparison of various methods, Computer Music Journal, Vol. 20, No. 2, summer 1996.

[Hourdin *et al.* 1997] C. Hourdin, G. Charbonneau, T. Moussa, A multidimensional scaling analysis of musical instruments' timevarying spectra. Computer Music Journal 19(1) 1997.

[Hummel *et al.* 1987] R. A. Hummel, B. Kimia, S. W. Zucker, Deblurring gaussian blur, Computer vision, graphics, and image processing 38, pp.66-80, 1987.

[IMA 1983] International MIDI Association, MIDI Musical Instrument Digital Interface Specification 1.0. 1983.

[ITU-R 85/10 1994] Recommendation ITU-R 85/10, "Methods for the subjective assessment of small impairments in audio systems, including multichannel sound systems". International Telecommunication Union, Geneva, Switzerland. 16 March 1994.

[Iverson *et al.* 1993] P. Iverson, C. L. Krumhansl, Isolating the dynamic attributes of musical timbre. J. Acoust. Soc. Am. 94(5), November 1993.

[Jaffe *et al.* 1983] D. A. Jaffe, J. O. Smith, Extension of the Karplus-Strong plucked-string algorithm. Computer Music Journal, Vol. 7, No. 2, summer 1983.

[Jensen 1988] K. Jensen, Construction d'un clavier MIDI, Projet de fin d'étude, ENSEEIHT, Toulouse, France, 1988.

[Jensen 1989] K. Jensen, Evaluation et construction d'un synthetiseur musical. Rapport de fin d'étude DEA, ENSEEIHT, Toulouse, France, 1989.

[Jensen 1993] K. Jensen, The French addword, an entropy approach. IBM internal, Paris, 1993.

[Jensen 1996a] K. Jensen, The control of musical instruments, Proceedings of the Nordic Acoustical Meeting, pp 379-383, The Acoustical Society of Finland, Helsinki, Finland, 1996.

[Jensen 1996b] K. Jensen, The control mechanism of the violin, Proceedings of the Nordic Acoustical Meeting, pp 373-378, The Acoustical Society of Finland, Helsinki, Finland, 1996.

[Jensen 1998] K. Jensen, Spectral envelope modeling. Proc. DSAGM, pp 91-97, Dept. of Computer Science, University of Copenhagen, 1998.

[Johansen 1994] P. Johansen, On the classification of toppoints in scale space. J. Math. Imaging and Vision. 4, 57-67, 1994.

[Kameoka *et al.* 1969] A. Kameoka, M. Kuriyagawa, Consonance theory part II: Consonance of complex tones and its calculation method. J. Acoust. Soc. Am. Vol. 45, No. 6, 1969.

[Karjalainen *et al.* 1993] M. Karjalainen, V. Välimäki, Z. Jánosy, Towards high-quality sound synthesis of the guitar and string instruments. Proc. ICMC 1993.

[Kimia *et al.* 1993] B. B. Kimia, S. W. Zucker, Analytic inverse of discrete gaussian blur. Optical engineering, Vol. 32, No. 1, January 1993.

[Klatt 1980] D. H. Klatt, Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 67(3), March 1980.

[Kleczowski 1989] P. Kleczowski, Group Additive Synthesis. Computer Music Journal 13(1), 1989.

[Klingholz 1987] F. Klingholz, The measurement of the signal to noise ratio (SNR) in continuous speech. Speech Communication 6, 1987.

[Kostek *et al.* 1996] B. Kostek, A. Wieczorkowska, Study of parameter relations in musical instrument patterns, AES preprint 4173 (E-6), Presented at the 100th Audio Engineering Society convention, Copenhagen, May 11-14 1996.

[Krimphoff *et al.* 1994] J. Krimphoff, S. McAdams, S. Winsberg, Caractérisation du timbre des sons complexes. II Analyses acoustiques et quantification psychophysique. Journal de Physique IV, Colloque C5, Vol. 4. 1994.

[Kroon *et al.* 1990] P. Kroon, B. S. Atal, Pitch predictors with high temporal resolution. Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. pp. 661-664, 1990.

[Kronland-Martinet 1988] R. Kronland-Martinet, The wavelet transform for analysis, synthesis, and processing of speech and music sounds. Computer Music Journal, 12(4), 1988.

[Krumhansl 1989] C. L Krumhansl, Why is musical timbre so hard to understand. *in* Structure and perception of electroacoustic sound and music, S. Nielzén, O. Olsson, editors. Excepta Medica 846, Elsevier, Amsterdam 1989.

[Lancaster *et al.* 1986] P. Lancaster, K. Salkauskas, Curve and surface fitting: An introduction, Academic Press, 1986.

[Lent 1989] K. Lent, An efficient method for pitch shifting digitally sampled sounds. Computer Music Journal, Vol. 13, No. 4, winter 1989.

[Leonard 1959] J. A. Leonard, Tactual Reaction times: I. Quarterly Journal of Experimental Psychology. 11, pp.76-83, 1959.

[Lindeberg 1996] T. Lindeberg, Edge detection and ridge detection with automatic scale selection, CVAP Report, KTH, Stockholm, 1996.

[Lindsay *et al.* 1977] P. H. Lindsay, D. A. Norman, Human information processing: An introduction to psychology. Academic Press, 1977.

[Mair *et al.* 1996] B. A. Mair, D. C. Wilton, Z. Réti, Deblurring the discrete gaussian blur. Proc. of the IEEE workshop MMBIA, San Francisco, June, 1996.

[Marques *et al.* 1994] J. S. Marques, A. J. Abrantes, Hybrid harmonic coding of speech at low bit-rates, Speech Communication 14, 1994.

[Mathworks 1992] Mathworks, Matlab reference guide. Mathworks Inc. Mass. 1992.

[Matthews *et al.* 1961] M. V. Matthews, J. E. Miller, E. E. David, Pitch synchronous analysis of voiced speech. J. Acoust. Soc. Am. Vol. 33, No. 2, February 1961.

[McAdams *et al.* 1995] S. McAdams, S. Winsberg, S. Donnadieu, G. de Soete, J. Krimphoff, Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research, 58, pp. 177-192. 1992.

[McAuley *et al.* 1986] R. J. McAulay, T. F. Quatieri, Speech analysis/synthesis based on a sinusoidal representation, IEEE Trans. on Acoustics, Speech and Signal Proc., vol. ASSP-34, No. 4, August 1986.

[McIntyre *et al.* 1981] M. E. McIntyre, R. T. Schumacher, J. Woodhouse, Aperiodicity in bowed-string motion, Acustica, Vol. 49, 1981.

[Medan *et al.* 1991] Y. Medan, E. Yair, D. Chazan, Super resolution pitch determination of speech signals. Proc. IEEE, 1991.

[Meddis *et al.* 1991a] R. Meddis, M. J. Hewitt, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. J. Acoust. Soc. Am. 89(6), June 1991.

[Meddis *et al.* 1991b] R. Meddis, M. J. Hewitt, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity. J. Acoust. Soc. Am. 89(6), June 1991.

[Mellody *et al.* 1997] M. Mellody, G. H. Wakefield, A model distribution study of the violin vibrato. Proc ICMC, 1997.

[Mitsuhashi 1982] Y. Mitsuhashi, Audio signal synthesis by functions of two variables. J. Audio Eng. Soc. 30(10), October 1982.

[Moore 1988] F. R. Moore, The dysfunction of MIDI, Computer Music Journal, Vol. 12, No. 1, Spring 1988.

[Moorer 1973] J.A. Moorer, The heterodyne filter as a tool for analysis of transient waveforms, Memo AIM-208, Stanford University, 1973.

[Moorer 1976] J. A. Moorer, The synthesis of complex audio spectra by means of discrete summation formulas. J. Audio. Eng. Soc. Vol. 24, No. 9, November 1976.

[Moré 1977] J. J. Moré, The Levenberg-Marquardt algorithm: Implementation and theory. Lecture notes in mathematics, Edited by G. A. Watson, Springer-Verlag, 1977.

[Møller 1973] AA. Møller, Basic mechanisms in hearing. Academic Press, New York and London, 1973.

[Møller 1996] A. Møller, Akustisk guitar syntese. Master Thesis, Computer Science Department, University of Copenhagen, 1996.

[Møller 1997] A. Møller, Transformation mellem musikinstrumenter, Master study report 96-11-2. Computer Science Department, University of Copenhagen, 1997.

[Nielsen 1995] Lars Bramsløw Nielsen, Subjective assessment of audio codecs and bitrates for broadcast purposes. Danmark Radio, R&D- Radio. 1995.

[Noll 1967] A. M. Noll, Cepstrum pitch determination. J. Acoust. Soc Am. Vol. 41, February 1967.

[Oohashi *et al.* 1997] T. Oohashi, E. Nishina, N. Kawai, Y. Fuwamoto, R. Yagi, M. Morimoto, Physiological and psychological effects of high frequency components above the audible range - an approach to KANSEI information processing. Proc. KANSEI - The technology of emotions. A. Camurri, Editor. University of Genova, 1997.

[Opolko *et al.* 1988] F. Opolko, J. Wapnick, McGill University Master Samples, 555 Sherbrooke Street West, Montreal, Quebec, Canada H3A 1E3. 1988.

[Paterson 1987] R. Paterson, A pulse ribbon model of monaural phase perception. J. Acoust. Soc. Am. Vol. 82, No. 5, November 1987.

[Phillips *et al.* 1996] D. Phillips, A. Parvis, S. Johnson, Multirate additive synthesis. Proc. of the Int. Comp. Music Conf. 1996.

[Pielemeier *et al.* 1996] W. J. Pielemeier, G. H. Wakefield, M. H. Simoni, Time-frequency analysis of musical signals. Proc. of the IEEE, Vol. 84, No. 9. September 1996.

[Plomb *et al.* 1965] R. Plomb, W. J. M. Levelt, Tonal consonance and critical bandwidth. J. Acoust. Soc. Am. 38, 1965

[Plomb *et al.* 1969] R. Plomb, H. J. M. Steeneken, Effects of phase on the timbre of complex tones, J. Acoust. Soc. Am.Vol. 46, No. 2, 1969.

[de Poli 1984] G. de Poli, Sound synthesis by fractional waveshaping. J. Acoust. Soc. Am. 32(11), November, 1984.

[Pollard *et al.* 1982] H. F. Pollard, E. V. Jansson, A tristimulus method for the specification of musical timbre. Acustica, vol. 51. 1982.

[Poulsen 1996] T. Poulsen, Psykoakustiske Målemetoder. IAT, Version 3.0. Danish Technical University, 1996.

[Press *et al.* 1997] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical recipes in C: The art of scientific computing. Cambridge University Press, 1997.

[Quatieri *et al.* 1986] T. F. Quatieri, R. J. McAulay, Speech transformations based on a sinusoidal representation, IEEE Trans. on Acoustics, Speech and Signal Proc., vol. ASSP-34, No. 6, Dec. 1986.

[Quinn 1994] B. G. Quinn, Estimating frequency by interpolating using frequency coefficients. IEEE Trans. on Signal Processing, Vol. 42, No. 5, May 1994.

[Quirós *et al.* 1994] F. J. C. Quirós, P. F-C. Enríquez, Real-time loose-harmonic matching fundamental estimation for musical signals. Proc. IEEE, 1994

[Rabiner *et al.* 1976] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, C. A. McGonegal, A comparative performance study of several pitch detection algorithms. IEEE Trans. ASSP, Vol. ASSP-24, No. 5, October 1976.

[Rabiner 1977] L. R. Rabiner, On the use of autocorrelation analysis for pitch detection. IEEE Trans. ASSP, Vol. ASSP-25, No. 1, February 1977.

[Rabiner *et al.* 1978] L. R. Rabiner, R. W. Schafer, Digital processing of speech signals. Prentice-Hall, 1978.

[Rayleigh 1896] J. W. S. Rayleigh, The theory of sound. reprinted in 1945 by Dover Publications. MacMillan company 1896.

[Richard *et al.* 1993] G. Richard, C. d'Allesandro, S. Grau, Musical noises synthesis using random formant waveforms. Stockholm Music Acoustic Conference, Pub. of the royal Academy of Music, Stockholm, Sweden 1993.

[Richard 1994] G. Richard, Modélisation de la composante stochastique de la parole. Thèse de Doctorat en Sciences de l’université Paris XI. 1994.

[Richard *et al.* 1996] G. Richard, C. d'Allesandro, Analysis, Synthesis and modification of the speech aperiodic component, Speech Communication 19, 1996.

[Risset 1965] J-C. Risset, Computer study of trumpet tones. J. Acoust. Soc. Am. Vol. 38 p. 912 (abstract), 1965.

[Risset *et al.* 1982] J-C. Risset, D. L. Wessel, Exploration of timbre by analysis and synthesis, *in* The psychology of music, D. Deutsch, editor, Academic Press, New York 1982.

[Risset 1991] J-C. Risset, Timbre analysis by synthesis: Representation, imitation and variants for musical composition. *in* Representations of musical signals. G. de Poli, A. Piccialli, C. Roads, Editors. The MIT Press, London 1991.

[Roads 1988] C. Roads, Introduction to granular synthesis. Computer Music Journal, Vol. 12, No. 2, summer 1988.

[Robinson 1982] E. A. Robinson, A historical perspective of spectrum estimation. Proc. of the IEEE, Vol. 70, No. 9, 1982.1

[Rodet *et al.* 1987] X. Rodet, P. Depalle, G. Poiret, Speech analysis and synthesis methods based on spectral envelopes and voiced/unvoiced functions. European Conference on Speech Technology, Edinburgh, September 1987.

[Rodet *et al.* 1992] X. Rodet, P. Depalle, Spectral envelope and inverse FFT synthesis. 93rd AES Convention, San Francisco, October, 1992.

[Rodet *et al.* 1997] X. Rodet, A. Lefèvre, The diphone program: New synthesis methods and experience of musical use. Proc. ICMC, 1997.

[Rossing 1990] T. D. Rossing, The science of sounds, Addison-Wesley, 1990.

[Rovan *et al.* 1997] J. B. Rovan, M. M. Wanderley, S. Dubnov, P. Depalle, Instrumental gestural mapping strategies as expressivity determinants in computer music performance. Proc. KANSEI - The technology of emotions. A. Camurri, Editor. University of Genova, 1997.

[Sandell *et al.* 1995] G. J. Sandell, W. L. Martens, Perceptual evaluation of principal-component-based synthesis of musical timbres. J. Acoust. Soc. Am. Vol. 43, No. 12, December 1995.

[Sandell 1998] G. J. Sandell, Sharc timbre database. http://sparky.parmly.luc.edu/sharc, 1998

[Schaeffer 1966] P. Schaeffer, Traité des objets musicaux. Editions de Seuil, 1966.

[Scheirer *et al.* 1997] E. Scheirer, M. Slaney, Construction and evaluation of a robust multifeature speech/music discriminator. Proc. ICASSP, April 21-24, Munich, 1997.

[Schroeder *et al.* 1979] M. R. Schoeder, B. S. Atal, J. L. Hall, Optimizing digital speech coders by exploiting masking properties of the human ear. J. Acoust. Soc. Am. Vol. 66, No. 6, December 1979.

[Schumacher *et al.* 1990] R. T. Schumacher, C. Chafe, Characterization of aperiodicity in nearly periodic signals. J. Acoust. Soc. Am. Vol. 91 No. 1, January 1992.

[Schwarz 1989] H. R Schwarz, Numerical Analysis, Wiley, 1989.

[Sekey *et al.* 1984] A. Sekey, B. A. Hanson, Improved 1-bark bandwidth auditory filter. J. Acoust. Soc. Am. Vol. 75, No. 6, 1984.

[Serra *et al.* 1990] X. Serra, J. Smith, Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition, Computer Music Journal, vol. 14, No. 4, winter 1990.

[Serra *et al.* 1997] X. Serra, J. Bonada, P. Herrera, R. Loureiro, Integrating complementary spectral models in the design of a musical synthesizer. Proc. of the Int. Comp. Music Conf. 1997.

[Skovenborg 1997] E. Skovenborg, Classification of monophonic tones. unpublished, 1997.

[Small 1959] A. M. Small, Pure-tone masking. J. Acoust. Soc. Am. Vol. 31, No. 12, December 1959.

[Steiglitz 1996] K. Steiglitz, A digital signal processing primer, with application to digital audio and computer music. Addison-Wesley, 1996.

[Stilson *et al.* 1996] T. Stilson, J. Smith, Alias-free digital synthesis of classic analog waveforms. Proc. ICMC, 1996.

[Strawn 1980] J. Strawn, Approximations and syntactic analysis of amplitude and frequency functions for digital sound synthesis. Computer Music Journal, Vol. 4, No. 3, fall 1980.

[Strawn 1985] J. M. Strawn, Modeling musical transitions. PhD thesis, Report No. STAN-M-26. CCRMA, Dept. of Music, Stanford University, 1985.

[Strong *et al.* 1966] W. Strong, M. Clark, Synthesis of wind-instrument tones. J. Acoust. Soc. Am. Vol. 41, No. 1, 1967.

[Sundberg 1987] J. Sundberg, The science of the singing voice, Northern Illinois University Press 1987.

[Tellman *et al.* 1995] E. Tellman, L. Haken, B. Holloway, Timbre morphing of sounds with unequal numbers of features. J. Audio Eng. Soc. Vol. 43, No. 9, September 1995.

[Terhardt 1974] E. Terhardt, On the perception of periodic sound fluctuations (roughness). Acustica, 30, 1974.

[Truax 1994] B. Truax, Discovering inner complexity: Time shifting and transposition with a real-time granulation technique. Computer Music Journal. 18(2), summer 1994.

[Ystad *et al.* 1996] S. Ystad, P. Guillemain, R. Kronland-Martinet, Estimation of parameters corresponding to a propagative synthesis model through the analysis of real sounds. Proc ICMC, 1996.

[Ystad 1998] S. Ystad, Sound modeling using a combination of physical and signal models. Doktor ingeniøravhandling, NTNU, Norway. 1998.

[Veldhuis 1998] R. Veldhuis, A computationally efficient alternative for the Linjencrants-Fant model and its perceptual evaluation. J. Acoust. Soc. Am. 103(1), January 1998.

[Vertegaal *et al.* 1996] R. Verrtegaal, B. Eaglestone, A comparison of input devices in an ISEE direct timbre manipulation task. Interacting with Computers, Vol. 8, No. 1, pp. 13-30, 1996.

[Wanderley *et al.* 1998] M. M. Wanderley, M. Battier, P. Depalle, S. Dubnov, V. Hayward, F. Iovino, V. Larcher, M. Malt, P. Piorrot, J. B. Rovan, C. Vergez, Gestural research at IRCAM: A progress report. Proc. Journée Informatique Musicale, La Londe-les-Maures, 1998.

[Weickert 1998] J. Weickert, Coherence-enhancing diffusion of colour images. *to appear in* Image and Vision Computing, 1998.

[Weickert 1999] J. Weickert, Coherence-enhancing diffusion filtering, *to appear in* Int. J. of Comp. Vision, 1999.

[Wessel 1979] D. Wessel, Timbre space as musical control structure. Computer Music Journal, 1979.

[Wessel 1997] D. Wessel, personal communication, 1997.

[Wolfram 1996] S. Wolfram, The mathematica book. Cambridge university press, 1996.

[ZIPI 1994] The ZIPI music interface language, Computer Music Journal 18(4), 1994.

[Zwicker *et al.* 1957] E. Zwicker, G. Flottorp, S. S. Stevens, Critical band width in loudness summation. J. Acoust. Soc. Am. Vol. 20, No. 5, May, 1957.

15. Table of Figures

FIGURE 1.1. COMPLETE FLOW CHART OF ANALYSIS AND MODELING IN THIS WORK. 3
FIGURE 2.1. ADDITIVE PARAMETERS PLOT. THE X AXIS IS TIME IN MS, THE Y AXIS IS FREQUENCY IN HZ AND THE Z AXIS IS AMPLITUDE..... 15
FIGURE 3.1. FFT-BASED PEAK SEARCH FOR A PIANO SOUND. FOUND PEAKS ARE MARKED WITH A '+'. THE SOLID LINE BELOW THE PEAKS IS THE MASKING LINE..... 22
FIGURE 3.2. FFT CANDIDATES FOR THE PIANO SOUND. 23
FIGURE 3.3. FREQUENCY DIFFERENCES FOR THE PIANO BEFORE (TOP) AND AFTER (BOTTOM) FILTERING.25
FIGURE 3.4. FREQUENCY DIFFERENCES AFTER CLEANING FREQUENCIES. '+' INDICATES ORIGINAL FREQUENCIES, '*' ARE CHOSEN FROM SEVERAL CANDIDATES, AND 'O' ARE NEW INSERTED FREQUENCIES. 25
FIGURE 3.5. FREQUENCY DIVIDED BY THE PARTIAL INDEX WITH ESTIMATED STRETCHED CURVE FOR THE PIANO SOUND..... 26
FIGURE 3.6. SPECTROGRAM OF THE FLUTE MELODY..... 29
FIGURE 3.7. SPECTROGRAM OF VIOLA MELODY..... 29
FIGURE 3.8. MOVING FUNDAMENTAL FREQUENCY FOR THE FLUTE MELODY..... 29
FIGURE 3.9. MOVING FUNDAMENTAL FREQUENCY FOR THE VIOLA MELODY..... 29
FIGURE 3.10. INSTANTANEOUS FREQUENCY OF THE FLUTE MELODY..... 30
FIGURE 3.11. INSTANTANEOUS FREQUENCY OF THE VIOLA MELODY..... 30
FIGURE 3.12. INSTANTANEOUS FREQUENCY AND EXTRACTED CURVE FOR THE FLUTE MELODY..... 31
FIGURE 3.13. INSTANTANEOUS FREQUENCY AND EXTRACTED CURVE FOR THE VIOLA MELODY..... 31
FIGURE 4.1. FFT ANALYZED ADDITIVE PARAMETERS FOR THE VIOLA..... 37
FIGURE 4.2. FFT ANALYZED ADDITIVE PARAMETERS FOR THE TRUMPET..... 37
FIGURE 4.3. FFT ANALYZED ADDITIVE PARAMETERS FOR THE PIANO..... 37
FIGURE 4.4. FFT ANALYZED ADDITIVE PARAMETERS FOR THE FLUTE..... 37
FIGURE 4.5. ILLUSTRATION OF THE TIME / FREQUENCY WINDOW DISCRIMINATION. A SMALL TIME DOMAIN WINDOW YIELDS A LARGE FREQUENCY DOMAIN WINDOW, AND VICE VERSE. 38
FIGURE 4.6. PRINCIPLE OF THE LTF FILTER CONSTRUCTION. TIME DOMAIN (TOP) AND FREQUENCY DOMAIN (BOTTOM)..... 39
FIGURE 4.7. ZERO-ORDER FILTERS AND SIGNAL FFT FOR A FLUTE SOUND..... 41
FIGURE 4.8. REBOUNDS OF FILTER..... 42
FIGURE 4.9. DETAIL OF THE FFT OF FILTER (TOP), SIGNAL (MIDDLE) AND RESULT OF FILTERING FOR THE FIFTH PARTIAL OF THE PIANO SOUND. X AXIS IS FREQUENCY BINS, AND Y-AXIS IS AMPLITUDE. ORIGINAL (DOTTED) AND AFTER ELIMINATION OF REBOUNDS (SOLID)..... 43
FIGURE 4.10. THE 5 FIRST HARMONIC OVERTONES OF PIANO C4 SOUND..... 44
FIGURE 4.11. THE 5 FIRST HARMONIC OVERTONES OF PIANO C4 AFTER ELIMINATION OF REBOUNDS..... 44
FIGURE 4.12. TIME RESOLUTION FOR 4 TEST SIGNALS. FFT ANALYSIS IS 'O' AND LTF ANALYSIS IS '*'.46
FIGURE 4.13. AMPLITUDE ERROR (TOP) AND FREQUENCY ERROR (BOTTOM) FOR THE FFT ANALYSIS 'O' AND THE LTF ANALYSIS '*'. 46
FIGURE 4.14. LTF BASED ADDITIVE PARAMETERS FOR THE VIOLA..... 47
FIGURE 4.15. LTF BASED ADDITIVE PARAMETERS FOR THE PIANO..... 47
FIGURE 4.16. LTF BASED ADDITIVE PARAMETERS FOR THE TRUMPET..... 47
FIGURE 4.17. LTF BASED ADDITIVE PARAMETERS FOR THE FLUTE..... 47
FIGURE 5.1. ADSR ENVELOPE..... 52
FIGURE 5.2. LINE SEGMENT APPROXIMATION OF ENVELOPE..... 52
FIGURE 5.3. PERCENT TIMES FOR THE VIOLA, THE TRUMPET, THE PIANO AND THE FLUTE..... 55
FIGURE 5.4. AMPLITUDE CURVES FOR THE VIOLA, THE TRUMPET, THE PIANO AND THE FLUTE. 55
FIGURE 5.5. AMPLITUDE AND FIRST DERIVATIVE FOR THE SMOOTHED FUNDAMENTAL OF FOUR SOUNDS WITH ENVELOPE TIMES FOUND WITH THE SLOPE METHOD..... 57
FIGURE 5.6. ENVELOPE OF SMOOTHED TRUMPET FUNDAMENTAL AND FIRST THREE DERIVATIVES WITH THE SLOPE POINTS. 58
FIGURE 5.7. SLOPE POINTS IN DIFFERENT SMOOTHING OF THE TRUMPET FUNDAMENTAL. UNSMOOTHED (TOP) TO VERY SMOOTHED (BOTTOM) 58
FIGURE 5.8. SLOPE TIMES FOR THE VIOLA, THE TRUMPET, THE PIANO AND THE FLUTE. 58
FIGURE 5.9. ATTACK (TOP) AND RELEASE (BOTTOM) PERCENTS FOR THE FOUR INSTRUMENTS..... 59
FIGURE 5.10. DIFFERENT SLOPES FOR THE ENVELOPE CURVE GOING FROM 0 TO 1..... 61
FIGURE 5.11. POSSIBLE CURVE FORMS FOR THE ATTACK(TOP) AND RELEASE (BOTTOM). 61

FIGURE 5.12. CURVE FITTING FOR THE ATTACK OF THE TRUMPET FUNDAMENTAL. 62
FIGURE 5.13. CURVE FORM VALUES FOR THE SLOPE ANALYSIS. ATTACK (TOP), SUSTAIN (MIDDLE) AND RELEASE (BOTTOM). NOTICE THE DIFFERENT Y SCALE FOR THE SUSTAIN CURVE FORM. 62
FIGURE 5.14. ORIGINAL AND PERCENTS (TOP) AND SLOPE (BOTTOM) FUNDAMENTAL ENVELOPE FOR FOUR SOUNDS 63
FIGURE 5.15. VIOLA ADDITIVE PARAMETERS. ORIGINAL (LEFT), PERCENT-BASED (MIDDLE) AND SLOPE-BASED (RIGHT). 64
FIGURE 5.16. PIANO ADDITIVE PARAMETERS. ORIGINAL (LEFT), PERCENT-BASED (MIDDLE) AND SLOPE-BASED (RIGHT). 64
FIGURE 5.17. TRUMPET ADDITIVE PARAMETERS. ORIGINAL (LEFT), PERCENT-BASED (MIDDLE) AND SLOPE-BASED (RIGHT) 65
FIGURE 5.18. FLUTE ADDITIVE PARAMETERS. ORIGINAL (LEFT), PERCENT-BASED (MIDDLE) AND SLOPE-BASED (RIGHT). 65
FIGURE 6.1. SPECTRAL ENVELOPE FOR THE VIOLA, THE PIANO, THE TRUMPET AND THE FLUTE. 70
FIGURE 6.2. FREQUENCY DIVIDED BY THE PARTIAL INDEX FOR THE VIOLA, THE PIANO, THE TRUMPET AND THE FLUTE. 70
FIGURE 6.3. ENVELOPE (TOP) AND FIRST DERIVATIVE (BOTTOM) TIMES FOR THE FUNDAMENTAL OF THE PIANO. 71
FIGURE 6.4. ATTACK (TOP) AND RELEASE (BOTTOM) TIMES FOR THE VIOLA, THE PIANO, THE TRUMPET AND THE FLUTE. 72
FIGURE 6.5. ORIGINAL AND RECREATED FUNDAMENTAL ENVELOPE FOR THE VIOLA, THE PIANO, THE TRUMPET AND THE FLUTE 73
FIGURE 6.6. INFLUENCE OF THE STD OF THE SHIMMER FROM 0 (TOP), 0.01, 0.1, 0.3 AND 0.5 (BOTTOM). NO JITTER, FILTER COEFFICIENT OF THE SHIMMER IS -0.5 75
FIGURE 6.7. INFLUENCE OF THE FILTER COEFFICIENT OF THE SHIMMER, WITH STD 0.3, FROM 0 (TOP), -0.3, -0.7, -0.9, -0.99 (BOTTOM), NO JITTER. 75
FIGURE 6.8. INFLUENCE OF THE STD OF THE JITTER FROM 0 (TOP), 0.01, 0.1, 0.3 AND 0.5 (BOTTOM). NO SHIMMER, FILTER COEFFICIENT OF THE JITTER IS -0.5 76
FIGURE 6.9. INFLUENCE OF THE FILTER COEFFICIENT OF THE JITTER FROM 0 (TOP), -0.3, -0.7, -0.9, -0.99 (BOTTOM). NO SHIMMER, STD OF THE JITTER IS 0.1. 76
FIGURE 6.10. PARTIAL FREQUENCY NOISE (JITTER) PARAMETERS. STANDARD DEVIATION (TOP), FILTER COEFFICIENTS (MIDDLE) AND CORRELATION (BOTTOM) FOR THE VIOLA (LEFT), THE PIANO, THE TRUMPET AND THE FLUTE (RIGHT). 77
FIGURE 6.11. PARTIAL AMPLITUDE NOISE (SHIMMER) PARAMETERS FOR THE VIOLA (LEFT), THE PIANO, THE TRUMPET AND THE FLUTE (RIGHT). 77
FIGURE 6.12. NOISE ENVELOPES. ATTACK AND RELEASE (DASHED) AND SUSTAIN (SOLID). 78
FIGURE 6.13. COMPLETE HLA SET FOR THE VIOLA SOUND. 80
FIGURE 6.14. COMPLETE HLA SET FOR THE PIANO SOUND. 80
FIGURE 6.15. COMPLETE HLA SET FOR THE TRUMPET SOUND. 81
FIGURE 6.16. COMPLETE HLA SET FOR THE FLUTE SOUND. 81
FIGURE 6.17. ORIGINAL AND MDA RECREATED ADDITIVE PARAMETERS FOR THE VIOLA. 82
FIGURE 6.18. ORIGINAL AND MDA RECREATED ADDITIVE PARAMETERS FOR THE PIANO. 82
FIGURE 6.19. ORIGINAL AND MDA RECREATED ADDITIVE PARAMETERS FOR THE TRUMPET. 83
FIGURE 6.20. ORIGINAL AND MDA RECREATED ADDITIVE PARAMETERS FOR THE FLUTE. 83
FIGURE 7.1. SPECTRAL ENVELOPE FOR THE VIOLA, THE PIANO, THE TRUMPET AND THE FLUTE. 87
FIGURE 7.2. SPECTRAL ENVELOPE FOR THE VIOLA, THE PIANO, THE TRUMPET AND THE FLUTE WITH SIMPLE BRIGHTNESS MATCHED CURVE. THE BRIGHTNESS OF EACH SOUND IN MARKED WITH A '*' 89
FIGURE 7.3. TIME DOMAIN (TOP) AND FREQUENCY DOMAIN BRIGHTNESS FUNCTION. THE PARTIAL INDEX BRIGHTNESS IS SET TO 3.0 90
FIGURE 7.4. TIME DOMAIN BRIGHTNESS FUNCTION WITH VARIABLE BRIGHTNESS GOING FROM 1 TO 10. 90
FIGURE 7.5. RESULTING SPECTRUM OF THE BCF FOR 4 SIGNALS WITH FUNDAMENTAL 200 HZ AND SAMPLE RATE 32 KHZ. BRIGHTNESS 2 (TOP), 4, 8 AND 16 (BOTTOM). 91
FIGURE 7.6. TRISTIMULUS VALUES FOR FOUR SOUNDS 93
FIGURE 7.7. SPECTRAL ENVELOPE FOR FOUR DIFFERENT IRREGULARITIES, 0, 0.1, 0.4 AND 0.7. BRIGHTNESS=5, TRISTIMULUS 1=0.25, TRISTIMULUS 2=0.5, ODD=0.3 94
FIGURE 7.8. A(1) TO A(4) IN B_{RANGE} 97
FIGURE 7.9. THE IRREGULARITY) IN B_{RANGE}. 97
FIGURE 7.10. SYNTHETIC SPECTRAL ENVELOPES FOR THE VIOLA, THE PIANO, THE TRUMPET AND THE FLUTE. 98
FIGURE 7.11. TIME VARYING SPECTRAL ENVELOPE PARAMETERS FOR THE VIOLA. 99
FIGURE 7.12. TIME VARYING SPECTRAL ENVELOPE PARAMETERS FOR THE PIANO. 99

FIGURE 7.13. TIME VARYING SPECTRAL ENVELOPE PARAMETERS FOR THE TRUMPET. 100
FIGURE 7.14. TIME VARYING SPECTRAL ENVELOPE PARAMETERS FOR THE FLUTE. 100
FIGURE 7.15. TIME VARYING AMPLITUDE OF THE FOUR TEST SOUNDS. VIOLA (TOP), PIANO, TRUMPET AND FLUTE (BOTTOM). 100
FIGURE 7.16. ORIGINAL (TOP) AND SPECTRAL ENVELOPE MODEL (BOTTOM) RECREATED ADDITIVE PARAMETERS FOR 4 SOUNDS, VIOLA, PIANO, TRUMPET AND FLUTE. 101
FIGURE 7.17. ILLUSTRATIONS OF THE FORMANT SEARCH USING A LOW 'A' SOUND. TOP PLOT IS ORIGINAL (SOLID), SYNTHETIC (DASH-DOTTED), AND WITH FORMANTS (DOTTED) SPECTRAL ENVELOPE, BOTTOM PLOT IS FORMANTS ONLY. 103
FIGURE 8.1. ANALYZED FREQUENCY (SOLID), AND MDA MODEL FREQUENCY (DOTTED) FOR 4 INSTRUMENT SOUNDS, VIOLA, PIANO, TRUMPET AND FLUTE. 107
FIGURE 8.2. SPECTRAL ENVELOPE FOUR 4 MUSICAL INSTRUMENTS, WITH THE MDA MODEL SPECTRAL ENVELOPE (DOTTED). 109
FIGURE 8.3. ATTACK AND RELEASE TIMES FOR THE 4 SOUNDS, WITH THE MDA MODEL ENVELOPE TIMES (DOTTED). ATTACK (TOP) AND RELEASE (BOTTOM) 110
FIGURE 8.4. END OF ATTACK (TOP) AND START OF RELEASE (BOTTOM) PERCENTS FOR THE 4 SOUNDS WITH THE MDA MODEL PARAMETERS (DOTTED). 111
FIGURE 8.5. ATTACK AND RELEASE CURVE FORM FOR THE 4 SOUNDS WITH THE MDA MODEL PARAMETERS (DOTTED). 111
FIGURE 8.6. PIANO RELEASE PERCENTS (SOLID) WITH ALL PARTIALS MODEL(DASHDOTTED) AND 32 FIRST PARTIALS MODEL (DOTTED). 112
FIGURE 8.7. SUSTAIN SHIMMER PARAMETERS FOR THE 4 INSTRUMENTS WITH THE MDA VALUES (DOTTED). STANDARD DEVIATION (TOP), FILTER COEFFICIENT (MIDDLE) AND CORRELATION (BOTTOM). 113
FIGURE 8.8. SUSTAIN JITTER PARAMETERS FOR THE 4 INSTRUMENTS WITH THE MDA VALUES (DOTTED). STANDARD DEVIATION (TOP), FILTER COEFFICIENT (MIDDLE) AND CORRELATION (BOTTOM). 114
FIGURE 8.9. FLUTE SHIMMER (TOP) AND JITTER (BOTTOM) WITH EXP. MODEL (DOTTED) AND 2ND ORDER POLYNOMIAL MODEL (DASHDOTTED) 115
FIGURE 8.10. RECREATED HLA PARAMETERS FOR THE VIOLA. 118
FIGURE 8.11. RECREATED HLA PARAMETERS FOR THE PIANO. 118
FIGURE 8.12. RECREATED HLA PARAMETERS FOR THE TRUMPET 118
FIGURE 8.13. RECREATED HLA PARAMETERS FOR THE FLUTE 118
FIGURE 8.14. RECREATED HLA PARAMETERS OF THE VIOLA, WITH ERROR TERM. 119
FIGURE 8.15 RECREATED HLA PARAMETERS OF THE PIANO, WITH ERROR TERM. 119
FIGURE 8.16 RECREATED HLA PARAMETERS OF THE TRUMPET, WITH ERROR TERM. 120
FIGURE 8.17 RECREATED HLA PARAMETERS OF THE FLUTE, WITH ERROR TERM. 120
FIGURE 8.18. ADDITIVE PARAMETERS FOR THE VIOLA. ORIGINAL (LEFT) MDA WITHOUT ERROR (MIDDLE) AND MDA WITH ERROR (RIGHT). 120
FIGURE 8.19. ADDITIVE PARAMETERS FOR THE PIANO. ORIGINAL (LEFT) MDA WITHOUT ERROR (MIDDLE) AND MDA WITH ERROR (RIGHT). 121
FIGURE 8.20. ADDITIVE PARAMETERS FOR THE TRUMPET. ORIGINAL (LEFT) MDA WITHOUT ERROR (MIDDLE) AND MDA WITH ERROR (RIGHT). 121
FIGURE 8.21. ADDITIVE PARAMETERS FOR THE FLUTE. ORIGINAL (LEFT) MDA WITHOUT ERROR (MIDDLE) AND MDA WITH ERROR (RIGHT). 121
FIGURE 9.1. IDA FREQUENCY BANDS FOR DIFFERENT MUSICAL INSTRUMENTS (PICTURE TAKEN FROM [LINDSAY *ET AL*. 1977]). 126
FIGURE 9.2. SPECTRAL PARAMETERS FOR THE PIANO. 129
FIGURE 9.3. SPECTRAL PARAMETERS FOR THE VIOLIN 129
FIGURE 9.4. SPECTRAL PARAMETERS FOR THE CLARINET. 130
FIGURE 9.5. SPECTRAL PARAMETERS FOR THE FLUTE. 130
FIGURE 9.6. SPECTRAL PARAMETERS FOR THE SOPRANO. BRIGHTNESS (TOP LEFT), TRISTIMULUS (TOP RIGHT), ODD (BOTTOM LEFT) AND IRREGULARITY (BOTTOM RIGHT) 130
FIGURE 9.7. AMPLITUDE FOR THE 5 INSTRUMENTS. PIANO (SOLID), VIOLIN, (DOTTED), CLARINET, (DASHDOTTED), FLUTE, (DASHED) AND SOPRANO (+-SOLID). 130
FIGURE 9.8. PARTIAL INDEX BRIGHTNESS FOR 5 INSTRUMENTS. PIANO (SOLID), VIOLIN, (DOTTED), CLARINET, (DASHDOTTED), FLUTE, (DASHED) AND SOPRANO (+-SOLID). 131
FIGURE 9.9. ODD PLUS TRISTIMULUS 1 FOR THE FIVE INSTRUMENTS. PIANO (SOLID), VIOLIN, (DOTTED), CLARINET, (DASHDOTTED), FLUTE, (DASHED) AND SOPRANO (+-SOLID). 131
FIGURE 9.10. INHARMONICITY FOR THE 5 INSTRUMENTS. PIANO (SOLID), VIOLIN, (DOTTED), CLARINET, (DASHDOTTED), FLUTE, (DASHED) AND SOPRANO (+-SOLID). 132
FIGURE 9.11. ENVELOPE PARAMETERS FOR THE PIANO 133
FIGURE 9.12. ENVELOPE PARAMETERS FOR THE VIOLIN. 133

FIGURE 9.13. ENVELOPE PARAMETERS FOR THE CLARINET. 134
FIGURE 9.14. ENVELOPE PARAMETERS FOR THE FLUTE. 134
FIGURE 9.15. ENVELOPE PARAMETERS FOR THE SOPRANO. ATTACK (LEFT) AND RELEASE (RIGHT). TIME (TOP), PERCENTS (MIDDLE) AND CURVE FORM (BOTTOM) 134
FIGURE 9.16. SUSTAIN CURVE FORM (TOP) AND SUSTAIN LENGTH (BOTTOM) FOR THE 5 INSTRUMENTS. PIANO (SOLID), VIOLIN, (DOTTED), CLARINET, (DASHDOTTED), FLUTE, (DASHED) AND SOPRANO (+-SOLID).134
FIGURE 9.17. START CURVE FORM (TOP) AND START OF ATTACK PERCENTS (BOTTOM) FOR THE 5 INSTRUMENTS. PIANO (SOLID), VIOLIN, (DOTTED), CLARINET, (DASHDOTTED), FLUTE, (DASHED) AND SOPRANO (+-SOLID). 135
FIGURE 9.18. NOISE PARAMETERS FOR THE PIANO. 136
FIGURE 9.19. NOISE PARAMETERS FOR THE VIOLIN. 136
FIGURE 9.20. NOISE PARAMETERS FOR THE CLARINET. 136
FIGURE 9.21. NOISE PARAMETERS FOR THE FLUTE. 136
FIGURE 9.22. NOISE PARAMETERS FOR THE SOPRANO. SHIMMER (LEFT) AND JITTER (RIGHT). STANDARD DEVIATION (TOP), FILTER COEFFICIENT (MIDDLE) AND CORRELATION (BOTTOM). 137
FIGURE 9.23. ATTACK SHIMMER (TOP) AND JITTER (BOTTOM)) FOR THE 5 INSTRUMENTS. PIANO (SOLID), VIOLIN, (DOTTED), CLARINET, (DASHDOTTED), FLUTE, (DASHED) AND SOPRANO (+-SOLID). 137
FIGURE 9.24. SPECTRAL ENVELOPE PARAMETERS FOR THREE DIFFERENT LOUDNESSES FOR THE PIANO. ALL LOUDNESSES (SOLID), *PIANO* (DOTTED), *MEZZO FORTE* (DASHDOTTED) AND *FORTE* (DASHED). 139
FIGURE 9.25. IDA FUNDAMENTAL AMPLITUDE FOR THREE DIFFERENT LOUDNESSES FOR THE PIANO. ALL LOUDNESSES (SOLID), *PIANO* (DOTTED), *MEZZO FORTE* (DASHDOTTED) AND *FORTE* (DASHED). 139
FIGURE 9.26. INHARMONICITY FOR THREE DIFFERENT LOUDNESSES FOR THE PIANO. ALL LOUDNESSES (SOLID), *PIANO* (DOTTED), *MEZZO FORTE* (DASHDOTTED) AND *FORTE* (DASHED)....140
FIGURE 9.27. ENVELOPE PARAMETERS FOR THREE DIFFERENT LOUDNESSES FOR THE PIANO. ALL LOUDNESSES (SOLID), *PIANO* (DOTTED), *MEZZO FORTE* (DASHDOTTED) AND *FORTE* (DASHED)....140
FIGURE 9.28. SUSTAIN CURVE FORM VALUES (TOP) AND SUSTAIN LENGTH (BOTTOM) FOR THREE DIFFERENT LOUDNESSES FOR THE PIANO. ALL LOUDNESSES (SOLID), *PIANO* (DOTTED), *MEZZO FORTE* (DASHDOTTED) AND *FORTE* (DASHED). 141
FIGURE 9.29. NOISE PARAMETERS FOR THREE DIFFERENT LOUDNESSES FOR THE PIANO. ALL LOUDNESSES (SOLID), *PIANO* (DOTTED), *MEZZO FORTE* (DASHDOTTED) AND *FORTE* (DASHED)....142
FIGURE 9.30. ATTACK SHIMMER (TOP) AND JITTER (BOTTOM) STD. ALL LOUDNESSES (SOLID), *PIANO* (DOTTED), *MEZZO FORTE* (DASHDOTTED) AND *FORTE* (DASHED). 142
FIGURE 9.31. SPECTRAL ENVELOPE PARAMETERS FOR THE CLARINET WITH DIFFERENT TEMPI. TOTAL (SOLID), *ALLEGRO* (DOTTED) AND *MODERATO* (DASHDOTTED). 144
FIGURE 9.32. AMPLITUDE FOR THE CLARINET WITH DIFFERENT TEMPI. TOTAL (SOLID), *ALLEGRO* (DOTTED) AND *MODERATO* (DASHDOTTED). 144
FIGURE 9.33. ATTACK AND RELEASE ENVELOPE PARAMETERS FOR THE DIFFERENT TEMPI OF THE CLARINET. TOTAL (SOLID), *ALLEGRO* (DOTTED) AND *MODERATO* (DASHDOTTED)....145
FIGURE 9.34. SUSTAIN CURVE FORM (TOP) AND TIMES (BOTTOM) FOR THE DIFFERENT TEMPI OF THE CLARINET. TOTAL (SOLID), *ALLEGRO* (DOTTED) AND *MODERATO* (DASHDOTTED)....145
FIGURE 9.35. NOISE PARAMETERS FOR THE DIFFERENT TEMPI OF THE CLARINET. TOTAL (SOLID), *ALLEGRO* (DOTTED) AND MODERATO (DASHDOTTED). 146
FIGURE 9.36. SPECTRAL ENVELOPE PARAMETERS FOR THE DIFFERENT STYLES OF THE CELLO. COMPLETE CELLO SET (SOLID), *STACCATO* (DOTTED), *SPICCATO* (DASHDOTTED) AND DASHED (*LEGATO*). 147
FIGURE 9.37. AMPLITUDES FOR THE DIFFERENT STYLES OF THE CELLO. COMPLETE CELLO SET (SOLID), *STACCATO* (DOTTED), *SPICCATO* (DASHDOTTED) AND DASHED (*LEGATO*)....147
FIGURE 9.38. ENVELOPE PARAMETERS FOR THE DIFFERENT STYLES OF THE CELLO. COMPLETE CELLO SET (SOLID), *STACCATO* (DOTTED), *SPICCATO* (DASHDOTTED) AND DASHED (*LEGATO*)....148
FIGURE 9.39. SUSTAIN CURVE FORM (TOP) AND TIMES (BOTTOM) FOR THE DIFFERENT STYLES OF THE CELLO. COMPLETE CELLO SET (SOLID), *STACCATO* (DOTTED), *SPICCATO* (DASHDOTTED) AND DASHED (*LEGATO*). 149
FIGURE 9.40. START (TOP) AND END (BOTTOM) TIMES FOR THE DIFFERENT STYLES OF THE CELLO. COMPLETE CELLO SET (SOLID), *STACCATO* (DOTTED), *SPICCATO* (DASHDOTTED) AND DASHED (*LEGATO*). 149
FIGURE 9.41. NOISE PARAMETERS FOR THE DIFFERENT STYLES OF THE CELLO. 150
FIGURE 10.1. SPECTRAL ENVELOPE OF THE ORIGINAL (SOLID) AND MODIFIED (DOTTED) PIANO. 162
FIGURE 10.2. MEAN FREQUENCIES, DIVIDED BY THE PARTIAL INDEX, OF THE ORIGINAL (SOLID) AND MODIFIED (DOTTED) PIANO. 163
FIGURE 10.3. ORIGINAL (TOP), TARGET, AND MODIFIED (BOTTOM) FUNDAMENTAL ENVELOPE AFTER TIME MODIFICATION. 164
FIGURE 10.4. ORIGINAL (TOP), TARGET, AND MODIFIED (BOTTOM) FUNDAMENTAL ENVELOPE AFTER THE PERCENTS MODIFICATION. 165

FIGURE 10.5. ORIGINAL (TOP), TARGET, AND MODIFIED FUNDAMENTAL ENVELOPE AFTER THE CURVE FORM MODIFICATION. 166
FIGURE 10.6. FREQUENCY MAGNITUDE RESPONSE (TOP) AND TIME SIGNAL (BOTTOM) FOR SUSTAIN SHIMMER OF THE FUNDAMENTAL OF THE PIANO, ORIGINAL (SOLID) AND MODIFIED (DOTTED). 168
FIGURE 10.7. ORIGINAL (TOP), CLEAN TRUMPET ENVELOPE, AND MODIFIED PIANO FUNDAMENTAL ENVELOPE AFTER THE SHIMMER MODIFICATION. 168
FIGURE 10.8. ORIGINAL AND MODIFIED (DOTTED) FUNDAMENTAL JITTER OF THE PIANO. FREQUENCY RESPONSE (TOP) AND TIME DOMAIN (BOTTOM). 169
FIGURE 10.9. ORIGINAL PIANO FUNDAMENTAL FREQUENCY (TOP), CLEAN TRUMPET FREQUENCY (MIDDLE) AND MODIFIED PIANO FREQUENCY (BOTTOM). THE FREQUENCIES HAVE BEEN OFFSET TO FACILITATE READING. 169
FIGURE 10.10. MODIFIED PIANO HIGH LEVEL ATTRIBUTES 170
FIGURE 10.11. ORIGINAL TRUMPET HIGH LEVEL ATTRIBUTES. 170
FIGURE 10.12. ADDITIVE PARAMETERS FOR THE PIANO (LEFT), THE MODIFIED PIANO(MIDDLE) AND THE TRUMPET (RIGHT). 171
FIGURE 11.1. FREQUENCY RANGE OF INSTRUMENTS. 175
FIGURE 11.2. TIMBRE ATTRIBUTES USED IN THE CLASSIFICATION. 177
FIGURE 11.3. FREQUENCY BRIGHTNESS FOR FIVE INSTRUMENTS, PIANO, VIOLIN, CLARINET, FLUTE AND SOPRANO. X AXIS IS SOUND INDEX AND Y AXIS IS FREQUENCY. 178
FIGURE 11.4. IDEAL SPECTRAL ENVELOPE PLOTTED UP TO NYQUIST FOR FIVE INSTRUMENTS, PIANO, VIOLIN, CLARINET, FLUTE AND SOPRANO. X AXIS IS SOUND INDEX AND Y AXIS IS AMPLITUDE. 179
FIGURE 11.5. AMPLITUDES AT NYQUIST FOR FIVE INSTRUMENTS, PIANO, VIOLIN, CLARINET, FLUTE AND SOPRANO. X AXIS IS SOUND INDEX AND Y AXIS IS AMPLITUDE. 179
FIGURE 11.6. SORTED EIGENVALUES FOR THE MDA TIMBRE ATTRIBUTES. 180
FIGURE 11.7. THE 5 INSTRUMENTS IN THE 3 FIRST PCA DIMENSIONS. THE PIANO IS '*', THE VIOLIN IS '+', THE CLARINET IS 'O', THE FLUTE IS '.' AND THE SOPRANO IS 'X'. OBSERVE THAT THE DATA IS ALSO PLOTTED ON THE X, Y AND Z PLANE. THE UPPER CLUSTER IN THE MIDDLE IS THE 3D PLOT. 181
FIGURE 11.8. EIGENVALUES FOR THE FIRST 3 PCA DIMENSIONS. 181
FIGURE 12.1. MEAN DEGRADATION AND 95 % CONFIDENCE INTERVAL FOR THE 5 MODELS FOR ALL INSTRUMENTS AND SUBJECTS. 191
FIGURE 12.2. MEAN DEGRADATION FOR THE FIVE INSTRUMENTS FOR ALL MODELS AND ALL SUBJECTS. 192
FIGURE 12.3. MEAN DEGRADATION FOR THE SUBJECTS FOR ALL MODELS AND ALL INSTRUMENTS. 192
FIGURE 12.4. MEAN DEGRADATION FOR THE 5 INSTRUMENTS FOR AL SUBJECT AND MODEL 2 (A/S). .. 193
FIGURE 12.5. MEAN DEGRADATION FOR THE PIANO SOUND AND MODEL 2 (A/S) FOR ALL SUBJECTS, AS A FUNCTION OF FUNDAMENTAL FREQUENCY. 193
FIGURE 12.6. MEAN DEGRADATION FOR THE HLA MODEL, ALL SUBJECTS, AS A FUNCTION OF INSTRUMENT. 194
FIGURE 12.7. MEAN DEGRADATION FOR THE MDA MODEL, ALL SUBJECTS, AS A FUNCTION OF INSTRUMENT. 195
FIGURE 12.8. MEAN DEGRADATION FOR THE IDA MODEL, ALL SUBJECTS, AS A FUNCTION OF INSTRUMENT. 195
FIGURE 12.9. MEAN DEGRADATION FOR THE 5 MODELS FOR ALL INSTRUMENTS AND SUBJECTS, WITH THE SOPRANO REMOVED. 196
FIGURE 12.10. COMPLETE SCORES FOR THE 75 SOUNDS AND THE 5 MODELS. 196

A. Sound Recordings

A.1. Violin

- Material.
  - Violin
  - Microphone Sony ECM 909, placed at ca 1m.
  - DAT Denon DTR-80P
  - Room normal furnished, ca, 5*3 m, H=2.5m.
- All recordings played by Elisa Andersen.
- A normal scale is played in two register, high (treble) or low (bas) notes
- Executions with different
  - tempo (fast, slow)
  - style (legato, spiccato, staccato)
  - intensity (piano, mezzo-forte, forte)
- filename: style-intensity-tempo.aiff
- Parameter varying executions,
  - Bow Flat angle (2 times)
  - Bow Long angle
  - Bow Position (4 positions)
  - Bow Pressure (3 pressures)
  - Bow Speed (3 speeds)
  - Vibrato Speed (3 times)
  - Vibrato Extend (3 times)
  - Normal to spring staccato
- filename: parameter-note.aiff

A.2. Viola

- Material.
  - Viola
  - Microphone Sony ECM 909 placed at ca 1m.
  - DAT Denon DTR-80P
  - Room normal furnished, ca, 5*3 m, H=3m.
- All recordings played by Klaus Hansen.
- Executions with different
  - tempo (fast, slow)
  - style (col_legno, con_sordino, detache, flautando, legato, martele, pizzicato, spiccato, staccato, sul_ponticello, sul_tasto)
- filename: style-tempo.aiff
- Parameter varying executions,
  - Bow Position, Bow Direction, Bow Elasticity
  - Bow Flat angle, Bow Long angle
  - Bow Force, Bow Speed
  - flageolet, glissando
  - Left Finger Timing, Silencing, Strings
  - Tremolo, Vibrato Speed, Vibrato Extend
  - Viola Angle, Viola Direction, Viola Position, Viola Slope
- filename: parameter.aiff

A.3. Cello

- Material.
  - Cello
  - Microphone Sony ECM 909 placed at ca 1m.
  - DAT Denon DTR-80P
  - Room, heavy furnished, ca, 5*3 m, H=2m.
- All recordings played by Dan Tørning.
- a simple scale is played in different,
  - tempo (fast, slow)
  - style (legato, spiccato, staccato)
  - intensity (pianissimo, fortissimo)
  - range (bas, mid, treble)
- filename: style-tempo-range.aiff (intensity is mezzo-forte)
  legato-intensity.aiff (slow, wide range
- Parameter varying executions, One single note per execution, bas or treble note
  - bow speed (3 speeds & 4 speeds)
  - bow pressure, (3 pressures)
  - bow long angle,
  - bow angle, (45 degree & 0 degree, twice)
  - bow elasticity,
  - vibrato speed,
  - vibrato extent,
  - attack (legato to spiccato),
- filename: Parameter-note.aiff

A.4. Saxophone

- Material.
  - tenor Saxophone Selmer Balanced Action 38, Otto Link #6 mouthpiece, pico royal 3 1/2 (worn)
  - Microphone Sony ECM 909 placed at ca 1m.
  - DAT Denon DTR-80P
  - Room: normal furnished, ca, 5*3 m, H=3m.
- All recordings played by Brian Thorsbro.
- A ‘C’ scale and a ‘C7’ accord is played in different executions
  - normal
  - soft (less blow force, lower jaw withdrawn)
  - hard (more blow force, less opening)
  - subtones (soft lower jaw, big mouth opening)
  - sing (sing into the moutpiece)
  - attack (with and without tongue)
- filename: execution-speed.aiff

A.5. Clarinet

- Material.
  - Clarinet Sib Noblet
  - Microphone AKG C410, placed at ca 1 m.
  - DAT Sony TCD-D10 Pro
  - Room dimensions. 6.2*4.2 height=3.4 (m)
- All recordings played by Richard Kronland-Martinet
- Executions with different
  - tempo (allegro, moderato)
  - style (legato, staccato, tenuto)
  - intensity (piano, mezzo-forte, forte)
  - Notes mi-sol-sib-do, and same intervals 1 octave and 1 fifth higher
  - filename: clar-intensity-tempo-style.aiff
  - crescendo, staccato and tenuto, mi2 and si4
- filename: clar-note-cresc[-ten].aiff

A.6. Flute

- Material.
  - Flute traversière en do ‘C’, ‘Mateki’
  - Microphone AKG C410, placed close to the mouth.
  - DAT Sony TCD-D10 Pro
  - Room dimensions. 6.2*4.2 height=3.4 (m)
- All recordings played by Sølvi Ystad
- Andante for flute, Mozart KV315 (extract)
- filename: flut-kv315.aiff
  - Executions with different
  - tempo (allegro, moderato)
  - style (legato, staccato, tenuto, detache)
  - intensity (piano, mezzo-forte, forte)
  - Notes do-mi-fa-sol-do, from do3 to do5
- filename: flut-intensity-tempo-style.aiff
  - crescendo, staccato and legato, sol3 and sol4 and do4
  - double-tongue and octaviation effects
  - vibrato in sol4
- filename: flut-note-cresc[-sta].aiff
  flut-octav.aiff
  flut-intensity-doubletongue.aiff

A.7. Soprano

- Material
  - Soprano Voice (lyrique)
  - Microphone, placed at ca 1m.
  - DAT.
  - Room normal furnished, ca, 5*3 m, H=3m.
- All recordings performed by Karin Andersen.
- A normal scale is performed with different vowels and consonants and intensitys.
- Executions with different
  - style (legato, tenuto)
  - intensity (piano, mezzo-forte, forte)
  - vowels (a-o-y)
  - consonant (none-k-b)
- filename: style-(consonant)vowel-intensity.aiff
- Furthermore, the high and low registers are performed in ‘a’ mf.
- filename: high_register.aiff
  low_register.aiff
- A crescendo was also performed.
- filename: crescendo.aiff

A.8. Piano

- Material.
  - Piano MIDI Yamaha Disklavier C6
  - Microphone KM84i (electrostatic, cardoid) placed 25 cm above the Mi3 string.
  - Preamp. Sonosax SX PR.
  - DAT Sony TCD-D10 Pro
  - Room dimensions. 6.2*4.2 height=3.4 (m)
  - No lock on the piano.
- All recordings done with MIDI in an isolated room
- Isolated sounds, each file one octave, twelve notes 400 ms long each.
- filename: oXoY.aiff, X is octave and Y is velocity, (40, 72 or 104)

B. Listening test instructions in danish

Oplæring

I oplæringsfasen skal lytteren lære at identificere og genkende forskellig forvrængning og forringelse som er skabt af det system der står under test. Når oplæringsfasen er overstået skal du vide hvad du lytter efter. Bagefter vil du blive spurgt om at blind-teste de samme slags lyde som du hører i oplæringsfasen. I oplæringsfasen skal du også lære test-proceduren.

Du vil høre både referensen (originalet) og den genskabte (komprimerede) lyd. Den første lyd er altid referensen og de næste to lyde er referensen og den komprimerede lyd i tilfældig rækkefølge. Du skal så vurdere forringelsen af de to sidste lyde i forhold til den første. Det er altså forringelsen mellem den første og den anden lyd, og forringelsen mellem den første og den tredje lyd der skal vurderes. Lydene er typiskt under et sekund lange og de kan høres igen, hvis nødvendigt. I lytteprøven skal du vurdere forringelsen i en skala fra 5.0 til 1.0.

	Forringelse
5.0	Ikke hørbar
4.0	Hørbar, men ikke generende
3.0	Lidt generende
2.0	Generende
1.0	Meget generende

Fordi en af de to lyde der står under test altid er den samme som referensen, skal et af de to vurderinger altid være 5. Hvis en af lydene lyder bedre end referensen, så betyder det at der er en ‘Hørbar, men ikke generende’ forskel og vurderingen bør ligge mellem 4.0 og 4.9.

Du bør tænke over hvordan du bedømmer de lydforringelser du hører i oplæringsfasen, men du bør ikke diskutere dette med andre forsøgspersoner.

Blind Forsøg

Formålet med blind-testen er at vurdere lyde af den type du kommer til at høre i oplæringsfasen.

I hver forsøg vil du høre 3 lyde, hvor den første altid er referensen, og de to næste er referensen og den komprimerede lyd i tilfældig rækkefølge. Du bliver ikke fortalt, hvilken af de to lyde der er referensen og hvilken der er den komprimerede lyd, derfor kaldes testen ‘blind’. Fordi en af lydene altid er den samme som originalen, skal en af vurderingerne altid være 5. Den anden vurdering skal gives i forhold til hvor brugbar lyden ville være i en almindelig musikalsk situation. Det betyder at hvis lyden lyder godt og som det samme instrument som originalet så skal den vurderes højt, også selvom den lyder anderledes end originalet. Der bør ikke laves vurdering for tonehøjdeforskelle, længdeforskelle, eller lydstyrkeforskelle.