Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing
Espoo 2003 Report 68

PERCEPTION OF ATTRIBUTES IN REAL AND SYNTHETIC STRING INSTRUMENT SOUNDS

Hanna Järveläinen

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Auditorium S4, Department of Electrical and Communications Engineering, Helsinki University of Technology, Espoo, Finland, on the 31st of January, 2003, at 12 o'clock noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Acoustics and Audio Signal Processing

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Akustiikan ja äänenkäsittelytekniikan laboratorio

HELSINKI UNIVERSITY OF TECHNOLOGY P.O. BOX 1000, FIN-02015 HUT http://www.hut.fi	ABSTRACT OF DOCTORAL DISSERTATION

Author Hanna Järveläinen

Name of the dissertation
Perception of Attributes in Real and Synthetic String Instrument Sounds

Date of manuscript January 20, 2003	Date of the dissertation January 31, 2003
☐ Monograph	☑ Article dissertation (summary + original articles)

Department	Electrical and Communications Engineering
Laboratory	Laboratory of Acoustics and Audio Signal Processing
Field of research	Psychoacoustics
Opponent(s)	Dr. Marc Leman
Supervisor (Instructor)	Prof. Matti Karjalainen

Abstract

This thesis explores the perceptual features of natural and synthetic string instrument sounds. The contributions are in formal listening experiments on a variety of features in musical sounds that have not been studied in detail previously. The effects of inharmonicity on timbre and pitch have been measured. The results indicate that the implementation of inharmonicity is not always necessary. The timbre effect is more salient in natural instruments, but for high tones a pitch difference may also be detected. Guidelines were given for compensation of the pitch effect. A perceptual study of the decaying parameters showed that large deviations from the reference value are tolerated perceptually. The studies on the audibility of initial pitch glides
and dual-polarization effects provides practical knowledge that helps in the implementation of these features in digital sound synthesis. Related to expression rather than basic string behavior, the study on perception-based control of the vibrato parameters has a sligthly different background. However, all of the studied features are more or less player-controlled by different ways of plucking the string or pressing the key.

The main objective of the thesis is to find answers to current problems in digital sound synthesis, such as parameter quantization. Another aim is to gain more general understanding of how we perceive musical sounds.

Keywords Perception, string instruments, sound source modeling, psychoacoustics

UDC 534.3:681.817:159.93	Number of pages 131
ISBN (printed) 951-22-6310-6	ISBN (pdf) 951-22-6314-9
ISBN (others)	ISSN 1456-6303

Publisher Helsinki University of Technology, laboratory of Acoustics and Audio Signal Processing

Print distribution Report 68 / HUT, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland

☑ The dissertation can be read at http://lib.hut.fi/Diss/

Preface

This work has been carried out in the Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland, from 1999 to 2002, and at the University of Padova, Italy, during my visit from November, 2001, to May, 2002. I have been a member in the Pythagoras graduate school and co-operated with Nokia Research Center.

I thank my supervisor, Prof. Matti Karjalainen, first of all for coming up with this research topic. Matti's relaxed but innovative attitude has guided me into the world of musical instrument sounds, which is a much nicer place than the tractor cockpit where I spent time during my Master's project. I admire Matti's patience – every time I postponed the defence date, he smiled. Or maybe he took it as humor.

Another important figure is Prof. Vesa Välimäki, who never saved his efforts in proof-reading, commenting, and motivating my work. His sincere enthusiasm definitely helped me complete this thesis. I also enjoyed working with Dr. Tero Tolonen and Dr. Tony Verma. Both being gurus in different aspects of sound synthesis, I got a broad picture of the field. Besides, Tero's humor and Tony's second-order humor added a lot to working life.

I am grateful to the pre-examiners of this thesis, Dr. Mikael Laurson of Sibelius Academy and Dr. Stephen McAdams of IRCAM, France, for their comments and critisism that helped improve the manuscript. I thank Prof. Giovanni De Poli of the University of Padova for his insightful comments on timbre perception and modeling, and Dr. Davide Rocchesso of the University of Verona for his efforts in bringing experimental psychology to my reach, as well as tutti i colleghi italiani for making my visit memorable.

The Acoustics laboratory is full of delightfully "pöljä" (translates as brilliant) people, who have helped me out in many ways. Ms. Lea Söderman has taken care of urgent, painful, and whatever practical issues. Mr. Jussi Hynninen, besides developing the GP2 subjective test system, has fixed my computer. I thank Dr. Cumhur Erkut for his practical help and discussions about sound synthesis, and express my gratitude and apologies to those who frequently volunteered in my listening experiments, Tomppa, Juha, Henkka, Paulo, and others (including Mara and his legendary sacrifice for science). Finally, I thank the members of Naistenhuone, Ms. Riitta Väänänen and Dr. Ville Pulkki, and Henkka the hang-around member for their support, friendship, good humor / bad humor, freshbites, and for replacing my computer with a Commodore 64.

I would like to thank my friends and the Dominante Choir for Something Completely Different, whether it means a budget flight over Novosibirsk or South African cows passing by the concert venue in a sand desert. Top moments! My special thanks go to my parents for always supporting me without asking too many times, what it actually was that they supported.

The financial support of Nokia Research Center and the Academy of Finland through the Pythagoras graduate school is gratefully aknowledged. Additional funding came from the European MOSART network and the Italian Ministry of foreign affairs.

Hanna Järveläinen Espoo, Finland, 8th January 2003

Table of Contents

List of Publications

This thesis summarizes the following articles and publications, referred to as [P1]-[P6]:

[P1] H. Järveläinen, V. Välimäki and M. Karjalainen. Audibility of the timbral effects of inharmonicity in stringed instrument tones. In *Acoustics Research Letters Online (ARLO)* 2(3):79-84, 2001.

[P2] H. Järveläinen and T. Tolonen. Perceptual tolerances for the decaying parameters in string instrument synthesis. In *Journal of the Audio Engineering Society*, 49(11), pages 1049–1059, 2001.

[P3] H. Järveläinen and V. Välimäki. Audibility of initial pitch glides in string instrument sounds. In *Proceedings of the International Computer Music Conference*, pages 282–285, Havana, Cuba, September 17–23, 2001.

[P4] H. Järveläinen, T. Verma and V. Välimäki. Perception and adjustment of pitch in inharmonic string instrument tones. *Journal of New Music Research*, 31(3), in press.

[P5] H. Järveläinen. Perception-based control of vibrato parameters in string instrument synthesis. Proc. *International Computer Music Conference*, pages 287–294, Gothenburg, Sweden, September 16–21, 2002.

[P6] H. Järveläinen and M. Karjalainen. Perception of beating and two-stage decay in dual-polarization string models. Proc. *International Symposium on Musical Acoustics*, Mexico City, December 9-13, 2002.

List of Symbols

a	loop filter coefficient
B	inharmonicity coefficient
β	modulation index
c	propagation velocity of transversal vibration
d'	(d prime), measure of sensitivity
d	diameter
F_t	string tension
f, f_0, f_n, f_s	frequency (in hertz), fundamental frequency
	frequency of nth partial and sampling frequency
g	loop gain
g_c	coupling parameter
$H(z)$	transfer function in the z-domain
L, L_I	real-valued and integral delay line length
λ	wavelength (in meters)
m	modulation depth
m_p, m_o	mixing parameters
ω	frequency (in radians per second)
ρ	mass density
σ^2	variance
T_0	fundamental period
τ	time constant
μ_N, μS	means of noise- and signal-induced excitation distributions

List of Abbreviations

AM	amplitude modulation
ANSI	American National Standards Institute
DSP	digital signal processing
ERB	equivalent rectangular bandwidth
FM	frequency modulation
JND	just noticeable difference
MM	mixed modulation
MPEG	Moving Picture Expert Group
ROC	receiver operating characteristics
SDT	signal detection theory
SMS	spectral modeling synthesis
2AFC	two-alternative forced choice

1. Introduction

1.1 Background

A variety of sound synthesis methods have emerged with the development of digital signal processing (DSP). They are utilized to create new timbres or imitate natural sounds, mostly of musical instruments. And indeed, they work so well that the average listener cannot distinguish a resynthesized sound from the original.

Multimedia and mobile communication systems released music from its cosy elevators, urging it to travel round the world much faster than the natural speed of sound – with low cost. The trend is still to use less data to transmit and store more high-quality content. In this race, sound synthesis methods have proven useful. For instance, the MPEG-4 multimedia standard [1] makes use of analysis/synthesis in coding of audio [2]. The current idea is not to transmit the actual sampled waveform of the sound, but to represent the sound in another form, as a much smaller amount of control data which is used to drive a synthesis model in the receiver end. The same quality of sound would be achieved by fewer bits.

When the waveform disappears, the traditional data reduction methods, such as the psychoacoustic model in MPEG-1 audio coding standard [3], become useless. To make the new scheme feasible and more efficient, we first need to find a complete parametric representation of sound to create the control data, and then reduce the control data somehow. It seems natural to analyze sound into its individual features, such as pitch, harmonicity, or vibrato, and then use as control parameters the features that are most salient perceptually. However, the last step is not completely taken, since perception used to represent "something completely different" for a significant part of the synthesis community.

1.1.1 At the same time in another galaxy . . .

Perception in this context means hearing. The structure and basic functioning of our hearing system is of course known; in fact, the motivation for most perceptual studies has been to reveal the secrets of the hearing mechanism. The research has produced valuable information for instance on the perception of high, low, pulsed, masked, or modulated pure tones, on the importance of periodicity in pitch perception, and various details on directional hearing. However, in this form the results are hard to apply on musical and other natural sounds, which typically have a rich spectral and temporal content, and not much of the perceptual knowledge made its way to audio signal processing applications.

In recent years, an effort has been made to understand the perception of environmental events. Given that we use our hearing system to observe the physical world, the main interest is in studying how the perception of an event is related to the physical properties of the vibrating system that cause it. Many results in *ecological acoustics* suggest that we can extract surprisingly exact information from the physical event based on the corresponding auditory event. For example, perception of geometric form by sound has been reported for width-to-height ratios of struck bars in [4], for the length of wooden rods dropping on a surface in [5], and for the dimensions and shape of thin vibrating plates in [6]. The perception of acoustic source characteristics was studied using walking sounds in [7]. The pitch cues allowing to discriminate 3-D resonator shapes were analyzed subjectively and computationally in [8]. Studies on musical sounds include for example perceived mallet hardness for percussive sounds [9] and the perception of legato articulation on the piano [10]. All of these studies have tried to find physical correlates for the source properties perceived by sound. Of course, the aim is not to find the exact physical differences between male and female walking footsteps, but to show that from natural sound events we perceive physical properties directly rather than abstract timbral cues.

Apart from many experiments on timbre and consonance of tones and chords [11], [12], [13], [14], [15], musical tones have received relatively little attention. The perception of isolated musical instrument tones was perhaps seen less important than music cognition in general. The consequence is a lack of perceptual knowledge that would fit the practical needs of sound synthesis [16]. From the synthesis viewpoint, it is important to study what kind of features we hear in typical musical sounds, or actually, what we do not hear. Anything which remains inaudible could be left out, which would reduce the complexity of the synthesis models and the amount of control data, and would thus allow computational savings.

1.1.2 The more you synthesize, the more you save!

The reduction of data in sound synthesis by perceptual means is not a new idea, however. Already during the early years of digital sound synthesis it was recognized that understanding the perceptual effects of synthesized sounds was essential to the further development of computer music applications. Risset and Mathews [17] were able to reduce the data required for synthesis of the trumpet, for instance, by isolating the most salient physical property of the sound.

Most of the perceptual studies concerning data reduction are related to the spectral represenation of sound. Additive synthesis requires a lot of storage space, sometimes even more than the original sound sample, because each of the time-varying amplitudes and frequencies of the individual harmonics must be controlled. However, the high degree of correlation in the amplitude and frequency envelopes of the individual partials was very promising in terms of efficient data reduction.

Related work is reported in [18], [19], [20], and [21].

The starting point for data reduction is different in this thesis. The parameters whose perceptual effects are measured are not related to the spectral representation of the sound, but rather they control computational models that are used to simulate the functioning of natural instruments. The method is called physical modeling [22], [23], [24]. According to the selection of parameters, properties are added to and subtracted from the resulting sound. In different instruments and different playing styles, pitch and loudness conditions, a different set of parameters is needed.

1.2 Scope of the Thesis

This thesis explores the perceptual effects of musical sounds. The emphasis is on features that can be synthesized by physical modeling in the same form as they are observed in natural instruments. The main experimental approach is the definition of perceptual tolerance thresholds for changes in various physical parameters. The objective is to gain knowledge about the perception of musical sounds and also to provide simple perceptual guidelines with the aim of data reduction through efficient parameter selection and quantization in model-based sound synthesis systems.

The results of this thesis can also be applied without the aim of data reduction. A strong motivation of this research is the understanding of natural musical instruments and the construction and control of new, virtual ones. Efficient parametrization of instrument models opens new applications in musical and artistic contexts.

The scope of the thesis is restricted to isolated plucked and struck string instrument tones; considering chords and melodies or other instrument types would result in too much complexity at this initial stage. The reported experiments are related to the perception of inharmonicity, decay, vibrato, pitch glides, and dual-polarization effects. Even though these features by no means constitute a complete description of timbre, they do cause many of the most prominent perceptual effects in plucked and struck string tones. Furthermore, they have been implemented in many source models of string instruments, but knowledge of their perception has been missing. Consideration of instrument body resonances is excluded from the research topics, although it naturally has a major effect on timbre. The reason is that in Commuted Waveguide Synthesis (CWS) [25], [26], the effect of the instrument body can be included in the excitation signal and thus neither filter structure nor control parameters are needed for its implementation.

1.3 Contents of the Thesis

The thesis consists of six publications and an introduction. The introduction gives background to the research. The main points of the acoustics of stringed instruments are presented in Chapter 2. Chapter 3 gives a short introduction to model-based sound synthesis. Chapter 4 concerns human perception of musical sounds, and Chapter 5 overviews signal detection theory (SDT), the experimental method used in this thesis. A summary of the publications is presented in Chapter 6 and conclusions and future directions in Chapter 7.

2. Acoustics of string instruments

Plucked or struck string instruments, such as the guitar or the piano, share a number of features from the sound production mechanism even to the way in which the player can control the timbre of the tone. In addition, there are nonlinearities and nonidealities present in the system that produce special characteristics for different instruments. This chapter introduces sound production in plucked or struck string instruments as well as the physical origin of the features whose perception is studied in the publications. These include decay, beats, two-stage decay, inharmonicity, pitch glides, and vibrato.

2.1 String motion

The basic string motion is common to all plucked or struck string instruments. The string is left to vibrate freely after it has been excited by plucking by the finger or striking by the hammer, for instance. The string vibration is a combination of the harmonic modes at multiple frequencies of the fundamental, which is the lowest mode with the longest wavelength. However, the harmonics whose nodes are at the point of the excitation, are not awakened, since the standing-wave minimum cannot coincide with the maximum of string displacement. For instance, if the string is plucked in the middle, the sound contains only the odd harmonics 1,3,5,..., who have maxima at the excitation point. However, this kind of ideal behavior is practically not encountered in real strings; the effect is much weaker.

The vibration is damped until the string is at rest again. Although some of the damping is due to the internal friction of the string, viscous dissipation in the air, and direct sound radiation of the string, the main cause is the coupling of the string vibrations to the soundboard through the bridge. In the guitar the string vibrations are transmitted to the top plate and then to the back plate and the air cavity. The resonances of the vibrating body boost certain frequencies, which very much contributes to the characteristic timbre of the instrument. In the guitar, the lowest resonance is around 100 Hz, and above 400 Hz the resonances become dense [27].

The piano is somewhat more complicated, since it has an enormous pitch and dynamic range. For the low keys, the mass of the strings has to be increased. This is usually done by wrapping them with copper instead of simply using heavier and stiffer unwrapped strings. The strings are struck by a hammer and the vibrations are transmitted to the soundboard through the bridge, but because of the great mass of the board, the coupling is too weak to produce a loud sound. For this reason, each key is connected to three strings that are slightly mistuned. At first, when the vibrations from the three strings are in phase, the energy is transferred rapidly to the soundboard. But soon after the "prompt sound" part the strings get out of phase, which results in a more slowly decaying "aftersound". The loud sound at the beginning is changed into a weaker but more sustained sound, and the tone has a two-part decay pattern [28].

2.2 Decay, beats, and two-stage decay

Normally, string vibrations decay exponentially in such a way that higher partials die out faster than lower ones. However, the two-stage decay can often be observed, even though only one string were excited. The phenomenon is due to polarization of the string vibration [28]. The string actually vibrates in three modes: the transversal, the longitudinal, and the torsional. The last two modes have relatively little importance for sound production in plucked string instruments, even though the longitudinal mode becomes significant in the low register of the piano [29]. The transversal mode is divided into horizontal and vertical components, which vibrate in the plane of the top plate and a plane perpenticular to it, respectively. The decay rates and the frequencies of the two polarizations are slightly unequal, which results in two-stage decay and slow beatings.

2.3 Inharmonicity

The frequencies of the partials of stringed instrument sounds are not exactly harmonic. This is caused by stiffness of real strings, which contributes to the restoring force of string displacement together with string tension. The strings are dispersive: the velocity of transversal wave propagation is dependent on frequency. If the string parameters are known, the frequencies of the stretched partials can be calculated in the following way [30]:

$$f_n = n f_0 \sqrt{1 + Bn^2} \tag{2.1}$$

$$B = \frac{\pi^3 Q d^4}{64 l^2 T} \tag{2.2}$$

In these equations n is the partial number, Q is Young's modulus, d is the diameter, l is the length and T is the tension of the string, and f_0 is the fundamental frequency of the string without stiffness. B is the inharmonicity coefficient for an unwrapped string. Its value depends on the type of string and the string parameters. Completely harmonic partial frequencies are obtained with $B = 0$. One should note that the frequency f_1 of the first partial in the inharmonic complex tone is actually higher than the fundamental frequency f_0 of the ideal string without stiffness. In addition to stiffness, there can be other sources of inharmonicity. For example, strong inharmonicity can be observed in the attack transients of harpsichord tones, where the restoring force of string displacement is nonlinear [31]. However, in this work the focus is on the systematic stretching of partials which is mainly caused by string stiffness.

One of the most evident effects of inharmonicity is the stretched tuning of the piano to maintain harmonic consonance between musical intervals [32]. Because of inharmonicity, the higher partials of low tones become sharp with respect to corresponding higher tones, and unpleasant beats occur. To minimize the beats, the bass range is tuned slightly flat and the treble range slightly sharp compared to the equal temperament. In the middle range, the adjustment is only a few cents (1/100 of a semitone), but it can be significant at both ends of the keyboard [33]. Lattard [34] has simulated the stretched tuning process computationally.

2.4 Initial pitch glides

The modulation of string tension is an important nonlinearity of a vibrating string. The string elongates with increasing displacement, reaching the maximum length twice during a vibration period. As a result the string tension also modulates with half the period of string vibration. The speed c of the transversal wave varies with string tension F_t according to [27]

$$c = \sqrt{\frac{F_t}{\rho}} \tag{2.3}$$

where ρ is the mass density of the string. The fundamental period of a linear string is given by

$$T_0 = \frac{\lambda}{c_{nom}} \tag{2.4}$$

where λ is twice the distance between string terminations. Replacing c_{nom} by the variable speed c suggests that the short time average time period is modulated in the same way as average string tension. Both attenuate towards the steady state value exponentially. The time constant is related to the time constant of overall attenuation of the vibration amplitude [35].

The perceptual effect of tension modulation is a rapid descent of pitch during the attack, which is characteristic to many plucked and struck string instruments in forte playing. It can be detected for instance in the clavichord [36], the guitar [37], and the kantele – a traditional Finnish string instrument [38]. In the clavichord, where string tension can be directly controlled by the player through key pressure, the effect is boosted by the mechanical aftertouch.

Fig. 2.1 shows a fundamental frequency estimate obtained from a recorded electric guitar tone by the autocorrelation method [37]. The f_0 estimate decreases exponentially with time from 499 to 496 Hz, giving a glide extent of approximately 3 Hz.

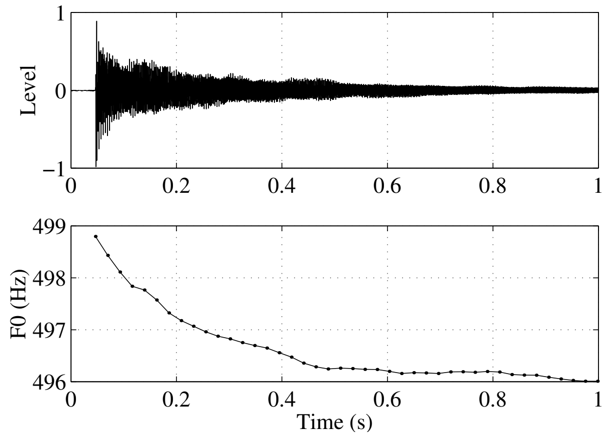

Figure 2.1: Waveform of a single tone played on the electric guitar (top) and its short-time fundamental frequency estimate, which shows a typical descent (bottom).

2.5 Vibrato

Vibrato is created by the motion of the player's finger back and forth on the finger board. The variable string length causes a constant frequency modulation. Vibrato is used because it gives the sound more depth. Another objective is to make the vibrato sounds stand out from the rest of the sound space. The origin and nature of vibrato in instrument sounds and especially voice is well-known [39], [40], [41], [42]. Research into the perception of vibrato concerns mainly emotional expression [43] or the pitch center of vibrato tones, which is subject to ongoing discussion [44], [45], [46].

Figure 2.2 presents the frequency modulation patterns analyzed from recorded classical guitar tones played by a professional guitar player [47]. The pitch of the tones is estimated by the autocorrelation method. It is seen that the modulation rate is typically around 5 Hz, while the total variation of pitch is between 0.7 Hz . . . 3 Hz.

Although the player creates mainly frequency modulation, it results in changes in amplitude that are crucial for the perception of vibrato [48], [49]. The moving harmonics are boosted and depressed according to the resonances of the instrument body. This poses problems for the systematic study of the perception of vibrato, since the body resonance characteristics vary from instrument to instrument, and the amplitude modulation changes for each note as a function of the depth of the frequency modulation. Mellody and Wakefield [49] showed that even though trig-

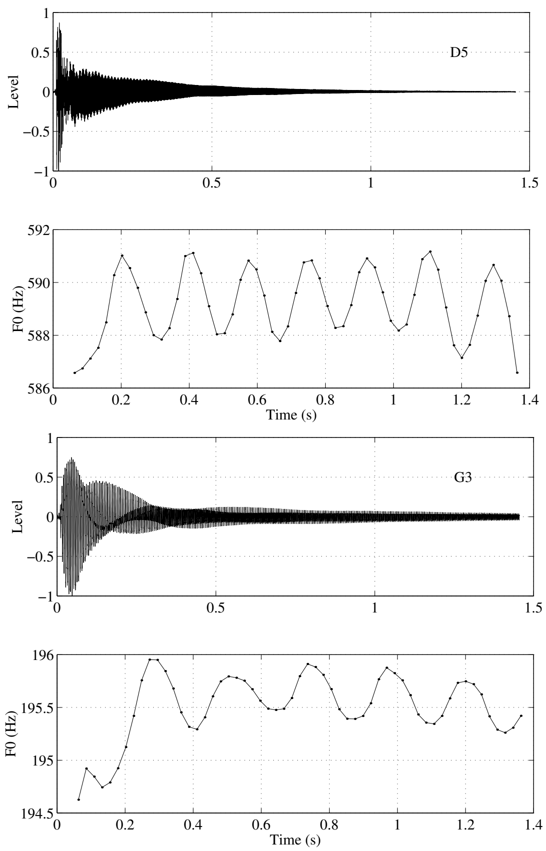

Figure 2.2: Waveform of a single vibrato tone played on the classical guitar (top), and a pitch estimate showing a typical frequency modulation pattern (bottom) – (a) D5, (b) G3. After [47].

gered by the sinusoidal frequency modulation, the amplitude changes were more complex in nature and the amplitude envelopes of individual harmonics had little or no correlation between each other.

2.6 Summary

This chapter has discussed the acoustics of string instruments. The general functioning is similar to all plucked or struck string instruments. Excitation of the string creates an exponentially damping vibration, which is coupled to the sound board through the bridge. The resonances of the instrument body boost certain frequencies and affect the resulting timbre.

The physical origins of the perceptual features studied in this thesis were discussed. Many of them are caused by nonlinearities and nonidealities of string vibration in real instrumnets. Polarization of the transversal vibration mode causes two-stage decay and beats. Stiffness makes a string dispersive, which results in elevation of the partials of their harmonic positions. String stiffness is the primary source of inharmonicity in the low register of the piano, for instance. The tension of the string is modulated due to its elongation with increasing displacement. As a consiquence, the attack is followed by the rapid descent in pitch. Vibrato is controlled by the player who moves his finger on the finger board, creating a nearly sinusoidal frequency modulation. As the moving harmonics coincide with different resonances of the instrument body, their amplitudes are modulated in a far more complex way.

3. Synthesis of string instrument sounds

Sound synthesis methods have developed greatly with the rise of digital signal processing. Pioneering work was published by Max Mathews as early as 1963 [50]. Since then, from abstract algorithms like FM synthesis [51], the interest has moved towards modeling of natural musical instruments [52]. The main categories are physical modeling, which models the sound source, and spectral modeling (or time-frequency modeling), which aims at constructing the spectrum that is received along the basilar membrane in the inner ear.

A physical model, or a sound source model, emulates computationally the generation and behavior of sound in natural musical instruments. Thus the sound produced by physical models is always connected to natural sound sources. The spectral modeling technique is based on splitting the spectral representation of the sound into its deterministic (sinusoidal) and stochastic (noise) components [53], [54]. This way the emulation of natural timbres is more elaborate, because the time-varying behavior of each partial has to be controlled individually. On the other hand, the method gains control over the spectral envelope, which is mostly responsible for the alterations in timbre.

Spectral modeling has found a broad field of applications because it is not restricted to any particular class of sounds, whereas a physical model is related to a specific sound source. However, physical modeling requires less memory and is efficient in producing many natural features of musical instruments, such as attack transients. Simulation of the physical properties of natural instruments enables studying the perception of sound source characteristics directly instead of abstract timbral cues. Since this is the main approach in the present studies, physical modeling formed a background for formulating the research problems and was primarily used for synthesizing the test tones.

3.1 Physical models of plucked strings

The vibrating string can be implemented by a one-dimensional digital waveguide [22], [23], [24], as seen in Fig. 3.1. The string has a transfer function

$$S(z) = \frac{1}{1 - z^{L_I} F(z) H_1(z)} \tag{3.1}$$

where L_I produces the integer part and $F(z)$ the fractional part of the delay line length. $H_1(z)$ is the loop filter which determines the decay of the tone by two parameters, g and a, according to

$$H_1(z) = \frac{g(1-a)}{1-az^{-1}} \tag{3.2}$$

For more natural sounding synthesis, a variety of features have been added. For instance, the model can be excited by a signal which includes the effect of the pluck as well as the instrument body [25], [26]. Other developments include fine-tuning of the pitch [22], [55], [56], dispersion simulation [57], [55], [58], [59], and various details of string vibration such as polarization and the plucking position [60] and tension modulation nonlinearity [37].

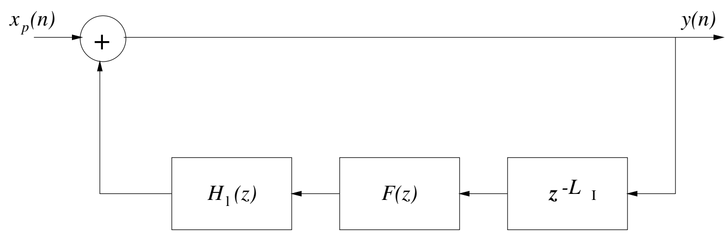

Figure 3.1: Block diagram of a simple string model [61].

3.2 Parametrization of the models

The demand for high-quality audio in a low-bitrate channel created the need for more parametric representations of sound. The MPEG-4 multimedia standard includes structured methods for representing synthetic audio [1], [62], [63]. Also in the more recent MPEG-7 multimedia content description interface [64], the timbre of musical sounds is described by a number of perceptually relevant parameters [65] (see [66] for an overview). The object-based approach offers a means to reduce the amount of data required for high-quality synthesis [67].

Sound objects are characterized in different ways depending on the synthesis method. When the spectral modeling technique is used, sound is parameterized while decomposed into its deterministic and stochastic parts. The method is widely used in audio coding [2] and analysis/synthesis of musical instruments.

In spectral modeling synthesis (SMS) [68], new sounds are synthesized by means of spectral domain transforms that affect the original analyzed parameters. Large parameter sets are usually needed, including the basic dynamic parameters such as fundamental frequency and amplitude envelope, and also higher level attributes like noisiness, harmonicity, vibrato, and spectral centroid. The behavior of the parameters changes over the different time segments of the sound: the attack, the steady state, and the release. Moreover, as timbre, pitch, and loudness effects in natural musical instruments are interconnected and related to the same physical attributes [69], the dynamic evolution of the higher level attributes should be modeled as a function of the basic parameters [70]. However, parameterization in the spectral domain makes many digital audio effects feasible [71].

In physical modeling, the quality of the sound is controlled by instrument-related parameters like pitch, decay rate, or plucking point, and the resulting timbre is a natural outcome of the model. The method agrees well with natural instruments. An important difference to spectral modeling is that the control parameters are used for simulating changes in the physical properties of the original instrument. This opens new possibilities to study the perception of source characteristics in musical instruments and, on the other hand, new questions about the quality assessment of synthesized sounds. It is essential to understand the perception of changes in the source characteristics and audibility of differences between the target and the synthesized sound.

Physical modeling also makes feasible higher-level mappings between the control of sound source models and complex musical information [72]. An attractive feature is also the possibility to create virtual instruments. An example of the "super guitar" with extended pitch range is given in [73].

For structured and object-based applications in audio, efficient schemes are needed for the selection and dynamic control of parameters. Perceptual studies can help us

- Understand the perceptual effects of individual parameters
- Select the most salient parameters
- Find a perception-based quantization and control schemes for the selected parameters

Earlier work concerning perceptual issues was discussed in Section 1.1.2. Until today, perception has received more attention in spectral modeling than in physical modeling, obviously because the spectral processing framework is closely related to other perceptual issues like timbre perception, recognition, and classification. However, the interest towards perceptual knowledge is increasing also in the physical modeling community.

3.3 Summary

This chapter gave a short introduction to the most widely used sound synthesis techniques: physical modeling and spectral modeling. The emphasis was on physical modeling, which enables studying the source characteristics of natural string instruments directly instead of abstract timbral ques. For this reason, the perceptual studies of this thesis are closely connected to the physical modeling framework.

A digital waveguide can be used to simulate the behavior of a string. The simple string model produces a damping, harmonic vibration. Additional effects encountered in real instruments, such as dispersion simulation, polarization, or tension modulation, can be modeled by different filter structures. Physical modeling also creates the possibility to control the source models by higher-level information related to the musical score or the player-instrument interaction.

The aim to control sound synthesis models effectively in parametric form has strongly increased the interest towards perceptual issues.

4. Perception of musical instrument tones

The previous chapter gave physical explanations for the typical features of string instrument sounds. The present chapter discusses the perceptual aspects of musical sounds on a general level, giving background to the perceptual studies presented in the publications.

The main perceptual features of musical sounds are pitch and timbre, which are both related to the processing of spectral components in the auditory system. The pitch of harmonic sounds is mainly determined by the fundamental frequency while timbre is created by the upper harmonics and their temporal envelopes. Thus changes in harmonicity or periodicity, caused by the physical properties of the vibrating string, induce changes also in pitch and timbre. The first sections present a summary of pitch and timbre perception and auditory organization, which form the essential background for the publications concerning pitch and timbre effects of inharmonicity.

Modulations are present in most musical tones in form of time-varying partial amplitudes and frequencies, vibrato, auditory beats, and pitch glides. Section 4.4 introduces the basic concepts of amplitude and frequency modulation and their perception. However, in natural instruments amplitude and frecuency modulations can hardly be separated from each other. They are present simultaneously, creating complex perceptual patterns that are hard to measure in analytic form. Thus the presentation in section 4.4 should be regarded as merely theoretical background to modulation detection.

4.1 Pitch perception

The American National Standards Institute (1973) states that pitch is ”that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low”. A pitch sensation can be weak or strong, and a single sound can cause many pitch percepts or none at all. Our hearing system can work both in analytical mode, which means that some of the pure tones included in a sound complex are heard separately, or in holistic mode, which means that we perceive one pitch as a joint effect of several components. The pitch sensation resulting from analytical listening is called spectral pitch, while holistic listening creates a virtual pitch sensation (also referred to as ”residue pitch” or ”low pitch”) [74]. Typically a sound or a group of simultaneous sounds evokes several spectral and virtual pitches. A pitch sensation can be related to noise as well as tonal sounds, but the pitches of narrowband or modulated noise or the repetition pitch of reflected sound sources is usually relatively weak [75]. The following gives an introduction to pitch perception of string instrument tones.

4.1.1 Pitch of pure and complex tones

The pitch of a pure tone usually corresponds to its frequency. Our hearing system can detect pitch differences as small as 0.1 . . . 0.2 % in the most sensitive frequency band between 1 kHz and 2 kHz [76], [77]. A great deal of musical sounds are periodic or quasi-periodic complex tones, which means that their spectrum consists of more or less harmonic line components. The harmonics fuse together into a single sound with a common timbre and pitch, which corresponds to the fundamental frequency of the complex. However, the presence of the fundamental is not necessary; the same pitch is perceived even though it is missing.

In many situations, pitch differs from frequency. The pitch of pure tones depends on their level. With increasing level, low tones (under 1 kHz) sound even lower and high tones (above 2 kHz) even higher than their frequency would suggest [78]. Also partial frequency masking can induce a remarkable pitch shift on pure tones [79]. The components of complex tones can be shifted from their harmonic positions. If the shift is towards lower frequencies, the separately heard inharmonic component seems to be lower than its frequency, while in upward shifting an even higher pitch is perceived [80], [81], [82], [83], [84]. In general, the pitch of complex harmonic tones is slightly lower than the fundamental suggests [85]. The intensity effect is also detected but more weakly than for pure tones [86].

Complex harmonic and inharmonic tones have unequal pitches. The effect of single mistuned components is relatively well known: up to 2 . . . 3 % mistuning, the change in residue pitch is approximately a linear function of the amount of mistuning. With increasing mistuning, the effect gets weaker and gradually disappears as the mistuned component segregates from the complex [87]. The ability of the partials to affect the residue pitch varies with partial number. There is some evidence that the first six harmonics are the most dominant [87] and that the dominance region drops towards lower partials with increasing fundamental frequency [88].

Pitch shifts of systematically mistuned stimuli have also been observed [89]. A pitch change is detected, when a harmonic tone is made inharmonic by changing either the center frequency [90] or frequency spacing of the components. The pitch difference was modelled in [91].

4.1.2 Models of pitch perception

The starting point for pitch perception is the basilar membrane in the cochlea, the inner ear cavity filled with fluids. The inner ear works as a frequency analyzer, whose resolution power for simultaneous components depends on their frequency. Low tones are resolved more accurately than high tones. The consequence is that only 4-6 of the lowest harmonics of a complex tone are resolved in the cochlea. The rest are analyzed as larger groups. The varying frequency resolution is described by the critical band scale [92] or the ERB (equivalent rectangular bandwidth) scale [93].

Theories have been proposed in two categories to explain the human pitch perception mechanisms. The frequency-place theories are based on the tonotopical coding of frequencies and the place of the maximum excitation along the basilar membrane. The place theory explains some features of pitch perception nicely, such as the effect of intensity. Pitch models, which are based on the tonotopic organization of the basilar membrane, propose that a spectral template is matched to the partials that are resolved in the cochlea. Models based on the place theory were proposed by Goldstein [94] and Terhardt [74], [78].

Evidence against the place theory is for instance the finding that also a group of unresolved partials can create a pitch percept [95]. Another group of theories and models originated from Licklider's idea of the timing of neural impulses [96]. The periodicity of the complex tone is contained in the interspike intervals of the nerve fibers for the higher, unresolved partials. Thus, according to the temporal theories, the ear works as a periodicity detector. The neural autocorrelation function can be used to detect common periodicity accross the auditory channels [97], [98], [99]. The autocorrelation model is able to extract information both from the resolved and unresolved partials, and it explains a major part of the phenomena in human pitch perception. Further evidence for the autocorrelation model was presented by the use of physiologically observed spike trains instead of simulated ones [100], [101].

4.2 Timbre perception

Timbre is defined by the American Standards Association [102] as "... that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are dissimilar". This obviously leaves lots of space for imagination. Sometimes timbre is referred to as "tone color". The further definition of timbre has been subject to speculations (see [69]). Frequent descriptions of timbre are *identity* and *tone quality*. Even though a sound maintains its identity under varying conditions, its quality may change in many ways. For instance, the voice of the same speaker sounds different heard acoustically from a short distance than heard over the telephone line.

Timbre is a combination of several perceptual dimensions. The main factors are related to the spectral content, the time-varying frequencies and amplitudes of the components. Also temporal events and modulations have a great effect on timbre. The attack transients of musical instrument sounds contribute to the characteristic timbre of each type of instrument so much that if they are absent, it is hard to recognize the instrument any more. It seems that the attacks are mostly important for the identification of sound sources, while the attributes that are present throughout the sound help us judge the timbre in a more general way [103].

4.2.1 Timbre space

There is an ongoing search for the perceptual and physical dimensions of timbre. The multidimensional scaling method (MDS) [104] has been utilized in many studies in order to find a relation between the perceived similarities of a set of timbres and a number of physical attributes. A low-dimensional space whose orthogonal dimensions represent the most prominent attributes is called the timbre space. A short distance between two timbres in the timbre space corresponds to high perceived similarity and a long distance to high dissimilarity.

The results obtained for musical instruments using the multidimensional scaling technique differ to a certain extent. The spectral centroid is a common finding [105], [106], [107], [108]. Others include synchronization of the transients and an attack-related attribute [105], steepness of the attack and the offset between the rise of the high and the low frequency harmonics [106], logarithmic rise time and the spectral flux [107], and logarithmic rise time and spectral irregularity [108]. These studies used almost entirely wind and string instruments. When the set of sounds was extended to percussive instruments [109], it was found that although two or three common dimensions could be found related to the spectral centroid and rise time, the results were generally context-dependent. This shows that a unitary timbre space is hardly likely to be found for all musical instruments. A new trend is to find physical correlates for the perception of various instrument classes instead of analytic perceptual dimensions like attack time or spectral centroid. In a study on struck bars [110], the MDS yielded a perceptual space that included also physical parameters such as material density.

4.3 Pitch, timbre, and auditory organization

The ear analyzes sounds into frequency bands, which can be modelled by the ERB scale as explained earlier. However, the information has to be put back together at some processing stage, since harmonic sounds obviously receive special treatment in the auditory system. As a result of spectral fusion, the partials of a complex tone are perceived as a single sound with common pitch and timbre. How does the ear decide, which partials belong to the same complex tone, and how can two simultaneous sounds with a simple pitch ratio be discriminated? It is assumed that the partials that fuse together are selected according to harmonic template matching. Also inharmonic partials can fuse successfully up to a certain degree of inharmonicity, after which they start to segregate from the complex. A number of features can reinforce the fusion. Common frequency or amplitude modulation or timing make the components fuse more easily. On the other hand, harmonicity is such a strong grouping criterion that even partials coming from different directions can fuse together [111].

4.4 Perception of modulations

Frequency and amplitude modulations have a strong role in musical instrument sounds, underlying such features as vibrato, auditory beats, and pitch glides. Moreover, the amplitudes and frequencies of the partials of musical tones vary considerably in time.

4.4.1 Amplitude modulation and beats

A sinusoidally amplitude modulated pure tone is presented as

$$x(t) = [1 + m\cos(\omega_m t + \phi)]\sin(\omega_c t) \tag{4.1}$$

where m is called the modulation depth. The amplitude of the carrier signal $\sin(\omega_c t)$ at frequency f_c is changed according to the modulating signal that fluctuates at frequency f_m. This results in a spectrum where in addition to the center frequency f_c there are two sidebands at frequencies $f_c - f_m$ and $f_c + f_m$ with amplitudes $m/2$. Beatings can be seen as balanced amplitude modulation, where there is no offset in the modulating signal. The spectrum consists only of the sum and difference frequencies while the center frequency is absent. The situation is similar to the summation of two sinusoids.

The detection thresholds for amplitude modulation, expressed as modulation percentage required for detection, have been measured in previous studies as a function of modulation frequency [112], [113], [114], being typically around $m = 0.05$ and decreasing for modulation rates higher than 64 Hz. The perception of amplitude modulation in the audible range varies with modulation frequency. When two sine tones are summed together, they are at first perceived as a single beating tone whose beating frequency equals the frequency separation of the tones. When the modulation rate increases over 10 ... 15 Hz, the loudness variation can no longer be followed and the beating is changed into a roughness sensation. With increasing frequency separation, the tones segregate from each other. When the frequency difference exceeds the critical bandwidth, the roughness sensation is lost as well [92].

If the amplitudes of the components of a beating tone are unequal, the overall pitch is varied according to the modulation rate. The maximum of the pitch shift cycle coincides with the minimum of the amplitude, so that the pitch effect is supressed. However, there are several reports that the pitch shift is sometimes detectable [115], [116].

4.4.2 Frequency modulation and mixed modulation

A sinusoidally frequency modulated (FM) signal is given by

$$x(t) = A\sin[\omega_c t + \beta\sin(\omega_m t + \phi)] \tag{4.2}$$

where $\omega_c = 2\pi f_c$ is a constant carrier frequency and $\beta = \Delta\omega/\omega_m$ is the modulation index, expressed as the ratio of the maximum excursion of the frequency variaton, $\Delta\omega$, and the modulation frequency, ω_m. If the modulation index is small, i.e., the frequency excursion is small compared to modulation frequency, the spectrum of a frequency modulated sine tone consists of the center frequency and two sidebands at frequencies $f_c + f_m$, $f_c - f_m$. The difference between amplitude modulation and narrow-band frequency modulation is a phase difference caused by the negative amplitude of the lower sideband in FM. For higher β we talk about wide-band FM. The spectrum of a wide-band FM signal includes an infinite number of sidebands, separated from the center frequency by its integer multiples.

The detection of frequency modulation, as well as the detection of amplitude modulation, depends on the modulation rate. Moreover, the detection tasks are almost identical because of the narrow-band simplification of the FM spectrum. The modulation index required for detection of FM decreases with modulation rate roughly from 0.7 for f_m = 4 Hz to 0.003 for $f_m = 64$ Hz [113].

It is hard to dicsriminate between frequency and amplitude modulation around the detection threshold [114]. There are different views about whether separate mechanisms exist for the detection of FM and AM [117], [118], [119], [113]. However, in instrument tones, as well as in other natural sounds, FM and AM seldom appear alone. Modulation where both FM and AM are present is called mixed modulation. If the modulations are in phase, the maximum loudness and maximum frequency excursion appear at the same time. If they are out of phase, the maximum loudness coincides with minimum frequency separation and vice versa. However, for sinusoids mixed modulation was more detectable than FM or AM alone, independent of the relative phase up to a modulation rate of 64 Hz [113].

4.4.3 Detection of glides

The perception of gliding tones is interesting, since variations of the fundamental frequency are often present in string instrument sounds. The detection task differs from normal frequency discrimination. In addition to discriminating the starting and end point frequencies, there is the question of detecting the gliding event itself. Previous research shows that glide detection is not entirely based on discrimination of the end points, but there is also sensitivity to glide extent and rate [120].

The detection thresholds of the glide extent are robust against changes in overall conditions. Neither center frequency, duration, level nor direction (upward or downward) have a strong effect on detectability. The detection thresholds are roughly a constant proportion of the ERB of the auditory filter. For instance, around the typical frequency region of musical instruments (below 2 kHz), a threshold of 0.15 ERB . . . 0.2 ERB was measured [121].

4.5 Summary

The present chapter explained the general basics of perception of musical sounds, forming important background to the specific studies presented in this thesis. The formation of pitch and timbre percepts was discussed. For harmonic tones, the fundamental frequency typically determines pitch, while the higher partials create timbre. However, many physical properties, such as inharmonicity, induce changes in pitch and timbre. Another important physical property, underlying such perceptual features as vibrato, beats, pitch glides, and musical consonance, is modulation. The concepts of frequency and amplitude modulation were explained and their perception was discussed on a general level.

5. Signal detection theory

The main objective of this work is finding perceptual detection or tolerance thresholds for various features of musical sounds. To be precise, no such things exist as thresholds. It can never be decided strictly when a signal or a difference between two signals is detected and when it is not. However, a simple measure is often needed for studying, how the detectability changes with a number of parameters. In such cases a threshold is usually decided to be somewhere between chance performance and 100-% performance. The thresholds used in this work were found by means of signal detection theory (SDT) [122].

The problem is similar to statistical hypothesis testing (Fig. 5.1). When there is no signal, the sensation is caused by noise alone and corresponds to the probability distribution on the left with mean μ_N. With signal present, the sensation caused by signal plus noise corresponds to the distribution on the right with mean μ_S. The observer's criterion is somewhere between the two means. Whenever the strength of the sensation exceeds the criterion, the observer judges that a signal is present and vice versa. Obviously, if the signal is strong, the distributions are well apart and the responses are mostly correct.

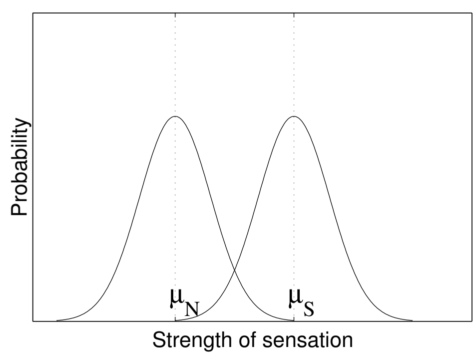

Figure 5.1: Probability density for noise (left) and signal + noise (right) excitations. Both distributions have the same variance σ^2.

A measure d', given by the distance between the two means, is used to describe the observer's sensitivity. Because μ_N and μ_S are based on hypothetical distributions, they remain unknown. By assuming Gaussian distributions underlying the decision-making, d' can be calculated from experimental data. Two experimental procedures are discussed in the following, the yes-no procedure and the two-alternative forced choice (2AFC) procedure.

In the yes-no procedure the observer is presented a stimulus which either does or does not contain the signal. The task is to detect the signal. In addition to correct detections and correct rejections, the answers can be false alarms, if a non-existent signal is detected, or misses, if an existing signal is not detected. After a number of trials, d' can be derived from the proportions of hits and false alarms. In the 2AFC method, the trials consist of two sound pairs, one of which contains the signal. The task is to decide, whether the signal occured in the first or the second interval. The measure of sensitivity can be directly derived from the proportion of correct responses.

The Receiver Operating Characteristics (ROC) curve plots the proportion of hits as a function of false alarms (Fig. 5.2). Isosensitivity curves are obtained by asking the listener to alter his or her decision behavior from "no" to "yes" responses while the stimulus intensity is kept constant. When the "no" answer is favored, there are few hits but also few false alarms. When the bias moves towards "yes" responses, the number of hits and false alarms increases. Thus the measurement produces an arch from point (0,0) to (1,1), all points of the curve having the same d'. Pure guessing, or chance performance, would produce a straight diagonal with $d' = 0$. With increasing sensitivity, the center of the arch would approach the upper left corner of the coordinate system.

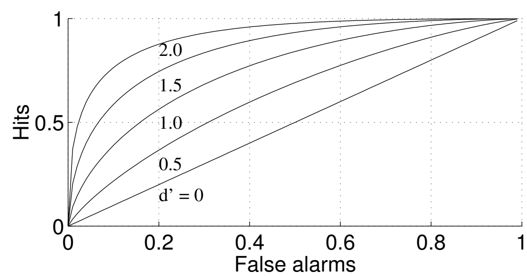

Figure 5.2: Receiver Operating Characteristics (ROC) curves for several values of d'.

In order to find a detection threshold, sensitivity measurements must be carried out for several stimulus intensities. The threshold can be expressed as stimulus intensity required for a given d' or area under the ROC curve. Commonly used thresholds, $d' = 1.0$ or 75 % area under the ROC curve, are also used in the present work. In the 2AFC method, the threshold can be expressed as the proportion of correct responses, for instance 75 %. It is also possible to use a staircase method that converges to the threshold value during the test session. The stimulus intensity is decreased after a correct response and increased after an incorrect one.

6. Summary of publications

The following presents a summary of the publications included in this thesis and describes the author's contribution. The initial idea for this research came from Professors Matti Karjalainen and Vesa Välimäki, who later contributed to many of the publications. Dr. Tero Tolonen and Dr. Tony Verma have also contributed as co-authors. In addition to them, the author has received ideas and practical help in sound synthesis from Dr. Cumhur Erkut, Mr. Henri Penttinen, and Mr. Janne Riionheimo. In general, the Sound source modeling team at Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, was always bursting with new questions about the perception of synthetic sounds, thus providing a natural framework of research topics, synthesis tools, and analysis data.

The scientific contributions of this thesis are not in sound synthesis methods but in the perceptual experiments, for which the author is mostly responsible alone. For this reason, the co-authors have not produced much text to the publications, but mostly ideas and synthesis tools. An exception is [P2], whose earlier version formed a part of Tero Tolonen's doctoral thesis.

Each publication in the thesis represents a different research topic. Together they form a study on the perception of synthesized string instrument sounds. None of the topics is completely covered in the publications; instead, it was considered worthwhile to address a wide area of questions concerning the perception of real and synthetic musical tones.

Publication [P1]

The first publication reports a study on the timbral effects of inharmonicity. The aim was to find out, how much inharmonicity is tolerated in plucked string instrument sounds until it affects the timbre. Earlier work on the perception of mistuned partials in complex tones and inharmonicity in musical tones is reviewed. In the present publication, audibility of inharmonicity in string instrument tones was measured in listening experiments for five fundamental frequencies and two tone durations. Inharmonicity was expressed in a way typical of dispersive strings, for which the frequencies of the higher partials can be calculated as a function of partial number and inharmonicity coefficient B. The test sounds included all partials up to 10 kHz.

Seven subjects participated in the experiment, where the task was to detect inharmonicity by judging two sounds, a harmonic sound an a possibly inharmonic sound, as same or different. The measurement was run for eigth values of B varied in equal steps within a range that was decided in initial listening. The detection threshold was expressed as B required for 75 % area under the ROC curve.

Inharmonicity was most detectable for the lowest note and least for the highest note: the mean threshold was more than 1,000 times higher for $C\sharp_6$ than for A_1. One reason for the results might be that low tones have simply more partials within the frequency band than high tones. A simple mathematical model was fitted to the data for estimating the audibility of inharmonicity as a function of fundamental frequency. When analyzed against available data on inharmonicity in the piano, it is noticed that high tones have a much higher B than low tones. Still the thresholds suggest that inharmonicity is clearly detectable in the bass range, while in the treble some of the measured inharmonicities reported in literature are lower than the mean detection thresholds. This indicates that because of its effect on timbre, inharmonicity should be implemented in the synthesis of low register piano tones, while for the high register the implementation is not always necessary.

The author designed and conducted the listening experiments as well as the analysis of the results. The test sounds were generated in co-operation with Vesa Välimäki.

Publication [P2]

Along with the pluck and the body response, decay is a crucial perceptual feature in plucked string instruments. Inaccurate reproduction of the decay pattern produces unnatural sounding synthesized tones. The main objective in this study was to measure the perceptual tolerance to variations in the decay parameters in plucked-string synthesis. The test sounds were generated by a physical model of the classical guitar, describing the decay by an overall time constant and a frequency-dependent parameter. The effects of the two parameters were investigated in separate tests. Two fundamental frequencies, tone durations, and types of excitations were used, resulting in eight sets of different sounds in both experiments. The decay parameter values were varied in nine equal steps around the reference values that were measured from a real guitar. In experiment one the overall time constant was the independent variable while the frequency-dependent parameter was fixed to the reference, and vice versa in experiment two.

Five normal-hearing subjects participated in the experiment. The task was to detect a difference between the test sound and the reference sound. The method of constant stimuli was used combined to the yes-no procedure. Each of the 72 test pairs in both experiments were judged 25 times. The experiments were conducted in five one-hour sessions. The first session was considered training.

In a perceptual sense, large variations are allowed for the decay parameters. The results indicate that values between 75 % and 140 % of the reference time constant and 83 % and 116 % of the frequency-dependent a parameter are perceptually transparent. Applied in sound synthesis using the measured reference parameter values, this means that the time constant can be varied within a region of about 300 ms and the -6-dB point of the frequency-dependent damping within a 300-Hz band.

The author was responsible for general implementation of the experiment, as well as statistical analysis of the results. The test signals were synthesized by Tero Tolonen, who also applied the results in model parameterization and parameter estimation. The test material was chosen in collaboration by both authors.

Publication [P3]

The publication studies the audibility of initial pitch glides in string instrument sounds due to modulation of string tension. Realistic sounding tones were generated by additive synthesis. The frequency decay rate of the glide was defined through the overall decay rate of amplitude, simulating the behavior in string instruments. The glide extent required for an audible pitch glide was measured for four fundamental frequencies in experiment 1 and as a function of decay time constant in experiment 2. The thresholds were expressed as 75 % correct responses, a measure derived from the proportions of hits and false alarms in the test trials. Five subjects participated in the experiments.

It was found that on the ERB frequency scale, the thresholds remained roughly constant at approximately 0.1 ERB with varying fundamental frequency. Thus, any pitch glide weaker than this could be left unimplemented in digital sound synthesis.

The test sounds were synthesized and the listening experiments were designed and conducted by the author.

Publication [P4]

This publication is closely related to publication [P1]. The effect of inharmonicity on pitch was measured by listening tests at five fundamental frequencies. Inharmonicity was defined in a way typical of string instruments, such as the piano, where all partials are elevated in frequency in a systematic manner. Six listeners participated in a pitch-matching test between inharmonic and harmonic tones. For each of the four fundamental frequencies, the pitch increase was measured as a function of B. The pitch judgment was usually dominated by one of the partials higher than the fundamental; however, with a high degree of inharmonicity, the fundamental became important as well.

Based on the results, guidelines are given for compensating for the pitch difference between harmonic and inharmonic tones in digital sound synthesis. The compensation is based on a pitch estimate given by the dominant partial. However, since inharmonicity reduces the pitch strength of sounds, upper limits are estimated for the relevance region of the compensation model. Beyond this region, the pitch judgment was usually ambiguous and individual partials became audible. In synthesis of natural string instrument sounds, the upper limits are practically never met, however.

The author's contribution is again designing and conducting the perceptual study. The synthesis tool for generating test sounds was created by Tony Verma, and it was based on sinusoidal modeling of the piano.

Publication [P5]

The paper reports four listening experiments exploring the perception of vibrato in string instrument synthesis. In model-based synthesis, vibrato is controlled by two parameters, rate and depth, which cause variation of the fundamental frequency. However, the most salient perceptual effect of vibrato is caused by amplitude variation which results from the resonances of the instrument body coinciding with the moving harmonics.

The perceived similarity of different vibrato patterns was studied by a rating experiment. Thereafter, JND's were measured for vibrato rate. Two additional experiments studied the effect of vibrato on the accuracy of pitch perception and the effect on musical consonance. Five subjects participated in the experiments. The results show that accurate control of the vibrato rate is much more important than control of vibrato extent. About 6 % of variation was allowed for vibrato rate, but for the extent the perceptual tolerances were much wider. Additional experiments on the perception of pitch and musical consonance of vibrato tones did not support the importance of vibrato extent either.

The author is alone responsible for the study. However, background information as well as help in analyzing the pitch variations in recorded sounds were given by Cumhur Erkut. Henri Penttinen provided the guitar body filter that was used to create the amplitude modulation effect to the test tones.

Publication [P6]

The presence of both vertical and horizontal polarizations creates auditory beats and a two-stage decay pattern. The perception of these effects was studied in listening experiments using an acoustic guitar string model.

Detecting a change in beating was measured as a function of the level difference between the vertical and horizontal components. Beating frequency and fundamental frequency were used as parameters. Perception of two-stage decay was studied in two experiments as a function of difference in the time constants of the polarization components. Perceptual tolerances were measured for variations in the time constant of the aftersound. The perceived similarity of one- and two-stage decay patterns was also studied. It was found that our sensitivity to dual-polarization effects is generally weak. Significant deviations from reference values were allowed before they became audible. The interaction of beating and two-stage decay phenomena should be considered when designing control schemes for synthesis parameters.

The author designed and conducted the experiments. A synthesis model for dual-polarization sounds was available at the Laboratory of Acoustics.

7. Conclusions

7.1 Summary

This thesis has its background in sound synthesis, where a number of questions have arisen concerning the perception of synthesized sounds and the effects of variations in the synthesis parameters. Perceptual studies have had an important role in the development of spectral modeling techniques since the early years; however, physical modeling has been left behind in this sense, and there is a lack of perceptual knowledge that fits its practical needs.

The aim of the thesis is to produce that kind of knowledge. The publications address different topics around synthesis of plucked or struck string instrument sounds. The tones were studied in isolation, out of the musical context, and without reverberation produced by the surrounding space. These constraints were necessary at this stage: musical sounds are complex enough without any extra variation.

The effects of inharmonicity on timbre and pitch are measured in publications [P1] and [P4]. The results indicate that the implementation of inharmonicity is not always necessary. The timbre effect is more salient in natural instruments, but for high tones a pitch difference may also be detected. Guidelines were given for compensation of the pitch effect.

Effects of variation in the decaying parameters were measured in [P2]. It was found that large deviations from the reference value are tolerated perceptually. Other basic features of string sounds were studied in [P3] and [P6], reporting measurements of the audibility of initial pitch glides and dual-polarization effects, respectively. Related to expression rather than basic string behavior, the study on perception-based control of the vibrato parameters in publication [P5] has a sligthly different starting point. However, all the studied features are more or less player-controlled by different ways of plucking the string or pressing the key.

The main objective of the thesis is to find answers to current problems in digital sound synthesis, such as parameter quantization. Another aim is to gain more general understanding of how we perceive musical sounds.

The separate studies form a set of perceptual tolerances and audibility thresholds of many aspects of string instrument tones. The common factor is that the features are selected according to and studied in a form related to physical modeling. Inharmonicity, decay, vibrato, pitch glides, and dual-polarization effects are essential features of plucked and struck string instrument tones and make a major contribution to the sound produced by a model. They can be implemented in the source models by special filter structures and controlled in parametric form. The current studies can help find out, which parameters must be controlled with the highest accuracy. For example, we perceive relatively small amounts of inharmonicity, small pitch glides, and small changes in vibrato rate, while much greater variations in decay and dual-polarization parameters are tolerated. Some of the results of this thesis are applied in the synthesis of clavichord tones in [123].

Knowledge of the perception of different source characteristics can also be utilized in synthesizing sounds that are based on natural instruments but cannot be produced by them. The perceptual guidelines can be used to decide, how much variation is needed in order to produce perceptually significant differences in synthetic sounds.

7.2 Future Directions

Several other features than the few that have been considered here deserve a closer perceptual study. Numerous details both in plucked and struck string instruments as well as other instrument groups contribute to their characteristic timbre. An important factor that is not included in this study is the instrument body. In sound source modeling the effect of the body can be included in the excitation signal to the model instead of a filter structure. However, knowledge of the perceptual effects of body resonances is needed to develope effective body models. Some research already exists, for instance on the timbral effects of the guitar body resonances [124] and models of reverberant instrument body responses [125].

The results from perceptual studies should be considered against theory and current auditory models. This would serve both as evaluation of the models and evaluation of the data from listening experiments. If the perception of musical sounds can be explained by basic models of our hearing system, also the perceptual effect of synthesis model parameters could be at least partly estimated without further listening experiments. It is hard to measure, however, how well the current results could be explained by auditory models until a full comparison can be done. No suitable models have been developed so far for some of the studied features, such as pitch glides or decay times. By intuition, the pitch variations caused by inharmonicity may well be explained by the various models of pitch perception. Also the perception of vibrato and beats produced by dual polarization may, at least partly, be explained by models and psychoacoustic indices of fluctuation strength or roughness.

The trend in multimedia, mobile communications and virtual audio is to find efficient structured and parametric representations of sound, which would enable transmitting high-quality content with a low bitrate. It has become more and more attractive to organize sounds as objects and control their features individually. Perceptual knowledge could be utilized in resynthesizing sounds whose properties are varying over time. The features included in the synthesis model could be reduced according to the perceptual thresholds, and the frequency of refreshing the control parameters could be minimized by the perceptual tolerances. For instance, the values of the vibrato parameters should be changed only if the effect is perceptually significant.

To build a working system like that, more perceptual studies on the features of musical sounds are needed in all instrument groups. Before those can be done, it is essential to isolate the physical features from natural sounds and turn them into synthesis parameters of instrument models. Much work is currently going on in automatic parameter estimation [126], [127], [128], [129], but more is needed in order to develope fast and reliable algorithms. More measurement data from natural instruments would also be welcome, so that the impact of the perceptual results could be better evaluated in relation to real instruments. However, now that perceptual methods are recognized as an important tool in assessing the quality of numerical models of musical instruments [130], fast development is expected in parallel with synthesis tool and parameter estimation techniques.

Physical modeling offers a gateway to higher-level parameterization. It seems natural to seek a relation between the synthesis parameters and musical information, as was already done in [72]. The next step is to go beyond musical notation and link the parametric representation of sound to models of expressivity and even emotions. Research into this field is presented in [131].

It should be noted, however, that the complexity of the perceptual effects increases, when musical context enters the picture. In this work only isolated tones are considered. The interaction between the partials of different tones in chords might either boost or supress the effects of some features. Modeling the evolution of the attributes according to the expressive intentions of the player is another wide field of study. Instead of complete answers, we have more and more open questions.

Bibliography

[1] ISO/IEC, "ISO/IEC IS 14496-3 Information Technology – Coding of Audiovisual Objects, Part 3: Audio," 1999.

[2] H. Purnhagen, B. Edler, and C. Ferekidis, "Object-based analysis/synthesis coder for very low bit rates." Preprint 4747, 104th Convention of the Audio Engineering Society, Amsterdam, May 1998.

[3] ISO/IEC, "ISO/IEC IS 11172-3 Information Technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio," 1993.

[4] S. Lakatos, S. McAdams, and R. Caussé, "The representation of auditory source characteristics: Simple geometric form," *Perception and Psychophysics*, vol. 59, pp. 1180–1190, 1997.

[5] C. Carello, K. Anderson, and A. Kunkler-Peck, "Perception of object length by sound," *Psychological Science*, vol. 9, no. 3, pp. 211–214, 1998.

[6] A. Kunkler-Peck and M. Turvey, "Hearing shape," *J. Experimental Psychology: Human Perception and Performance*, vol. 26, no. 1, pp. 279–294, 2000.

[7] X. Li, R. Logan, and R. Pastore, "Perception of acoustic source characteristics: Walking sounds," *J. Acoust. Soc. Am.*, vol. 90, no. 6, pp. 3036–3049, 1991.

[8] D. Rocchesso, "Acoustic cues for 3-D shape information," in *Proc. Int. Conf. Auditory Display, ICAD*, (Espoo, Finland), pp. 175–180, July 2001.

[9] D. Freed, "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events," *J. Acoust. Soc. Am.*, vol. 87, no. 1, pp. 311–322, 1990.

[10] B. Repp, "Acoustics, perception, and production of *legato* articulation on a digital piano," *J. Acoust. Soc. Am*, vol. 97, no. 6, pp. 3862–3874, 1995.

[11] R. Plomp, "Timbre as a multidimensional attribute of complex tones," in *Frequency analysis and periodicity detection in hearing* (R. Plomp and G. Smoorenburg, eds.), pp. 397–414, Sijthoff, Leiden, 1970.

[12] R. Plomp and W. Levelt, “Tonal consonance and critical bandwidth,” *J. Acoust. Soc. Am.*, vol. 38, pp. 548–560, 1965.

[13] A. Kameoka and M. Kuriyagawa, “Consonance theory part I: Consonance of dyads,” *J. Acoust. Soc. Am.*, vol. 45, pp. 1451–1459, 1969.

[14] A. Kameoka and M. Kuriyagawa, “Consonance theory part II: Consonance of complex tones and its calculation method,” *J. Acoust. Soc. Am.*, vol. 45, pp. 1460–1469, 1969.

[15] M. Leman, “Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (SIM),” in *Proc. COST G-6 Conference on Digital Audio Effects (DAFx-00)*, (Verona, Italy), pp. 1–6, Dec. 2000.

[16] J. Chowning, “Digital sound synthesis, acoustics, and perception: A rich intersection,” in *Proc. COST G-6 Conference on Digital Audio Effects (DAFX-00)*, (Verona, Italy), 2000.

[17] J.-C. Risset and M. Mathews, “Analysis of musical instrument tones,” *Physics Today*, vol. 22, no. 2, pp. 23–30, 1969.

[18] G. Charbonneau, “Timbre and the perceptual effects of three types of data reduction,” *Computer Music J.*, vol. 5, no. 2, pp. 10–19, 1981.

[19] P. Kleczkowski, “Group additive synthesis,” *Computer Music J.*, vol. 13, no. 1, pp. 12–20, 1989.

[20] G. Sandell and W. Martens, “Perceptual evaluation of principal-component-based synthesis of musical timbres,” *J. Audio Eng. Soc.*, vol. 43, no. 12, 1995.

[21] S. McAdams, J. W. Beauchamp, and S. Meneguzzi, “Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters,” *J. Acoust. Soc. Am.*, vol. 105, no. 2, pp. 882–897, 1999.

[22] D. A. Jaffe and J. Smith, “Extensions of the Karplus-Strong plucked-string algorithm,” *Computer Music J.*, vol. 7, no. 2, pp. 56–69, 1983.

[23] J. Smith, “Physical modeling using digital waveguides,” *Computer Music J.*, vol. 16, no. 4, pp. 74–91, 1992.

[24] J. Smith, “Principles of digital waveguide models of musical instruments,” in *Applications of Digital Signal Processing to Audio and Acoustics* (M. Kahrs and K. Brandenburg, eds.), ch. 10, pp. 417–466, Kluwer, 1998.

[25] J. Smith, “Efficient synthesis of stringed musical instruments,” in *Proc. Int. Computer Music Conf. ICMC’93*, (Tokyo, Japan), pp. 64–71, 1993.

[26] M. Karjalainen, V. Välimäki, and Z. Jánosy, "Towards high-quality sound synthesis of the guitar and string instruments," in *Proc. Int. Computer Music Conf.*, pp. 56–63, 1993.

[27] N. Fletcher and T. Rossing, *The Physics of Musical Instruments*. New York, USA: Springer-Verlag, 1991.

[28] G. Weinreich, "Coupled piano strings," *J. Acoust. Soc. Am.*, vol. 62, pp. 1474–1484, 1977.

[29] H. A. Conklin, "Design and tone in the mechanoacoustic piano. Part 3. Piano strings and scale design," *Journal of the Acoustical Society of America*, vol. 100, pp. 1286–1298, 1996.

[30] H. Fletcher, E. D. Blackham, and R. Stratton, "Quality of piano tones," *J. Acoust. Soc. Am.*, vol. 34, no. 6, pp. 749–761, 1962.

[31] W. Weyer, "Time-frequency-structures in the attack transients of piano and harpsichord sounds I, II," *Acustica*, vol. 36, pp. 241–258, 1976.

[32] H. Suzuki, "Effect of inharmonicity of a piano on tuning," in *Music and Concert Hall Acoustics: Conference Proceedings from McHa 1995*, pp. 219–232, Academic Press, 1996.

[33] D. Martin and W. Ward, "Subjective evaluation of musical scale temperament in pianos," *J. Acoust. Soc. Am*, vol. 33, pp. 582–585, 1961.

[34] J. Lattard, "Influence of inharmonicity on the tuning of a piano – measurements and mathematical simulation," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 46–53, 1993.

[35] K. Legge and N. Fletcher, "Nonlinear generation of missing modes on a vibrating string," *J. Acoust. Soc. Am.*, vol. 76, no. 1, pp. 5–12, 1984.

[36] V. Välimäki, M. Laurson, C. Erkut, and T. Tolonen, "Model-based synthesis of the clavichord," in *Proc. Int. Computer Music Conf.*, (Berlin, Germany), pp. 50–53, 2000. Full paper and sound examples are available at http://www.acoustics.hut.fi/publications/.

[37] T. Tolonen, V. Välimäki, and M. Karjalainen, "Modeling of tension modulation nonlinearity in plucked strings," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 3, pp. 300–310, 2000.

[38] V. Välimäki, M. Karjalainen, T. Tolonen, and C. Erkut, "Nonlinear modeling and synthesis of the kantele – a traditional Finnish string instrument," in *Proc. Int. Computer Music Conf.*, (Beijing, China), pp. 220–223, 1999. Full paper and sound examples are available at http://www.acoustics.hut.fi/publications/.

[39] H. Dejonckere, M. Hirano, and J. Sundberg, eds., *Vibrato*. London: Singular Publishing Group, 1995.

[40] E. Prame, "Measurements of the vibrato rate of ten singers," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 1979–1984, 1994.

[41] E. Prame, "Vibrato extent and intonation in professional western lyric singing," *J. Acoust. Soc. Am.*, vol. 102, no. 1, pp. 616–621, 1997.

[42] J. Gleiser and A. Friberg, "Vibrato rate and extent in violin performance," in *Proc. 6th Int. Conf. Music Perception and Cognition*, (Keele University, United Kingdom), 2000.

[43] S. Jansens, G. Bloothooft, and G. de Krom, "Perception and acoustics of emotions in singing," in *Proc. 5th Eurospeech*, vol. 4, (Rhodes, Greece), pp. 2155–2158, 1997.

[44] C. d'Alessandro and M. Castellengo, "The pitch of short-duration vibrato tones," *J. Acoust. Soc. Am.*, vol. 95, no. 3, pp. 1617–1630, 1994.

[45] J. Brown and K. Vaughn, "Pitch center of stringed instrument vibrato tones," *J. Acoust. Soc. Am.*, vol. 100, no. 3, pp. 1728–1735, 1996.

[46] L. Yoo, S. Moore, D. Sullivan, and I. Fujinaga, "The effect of vibrato on response time in determining the pitch relationship of violin tones," in *Proc. Int. Conf. Music Perception and Cognition*, (Seoul, South Korea), August 1998.

[47] C. Erkut, V. Välimäki, M. Karjalainen, and M. Laurson, "Extraction of physical and expressive parameters fot model-based sound synthesis of the classical guitar," in *Proc. 108th AES Convention*, 2000. Preprint 5114.

[48] M. Mathews and J. Kohut, "Electronic simulation of violin resonances," *J. Aoucst. Soc. Am.*, vol. 53, no. 6, pp. 1620–1626, 1973.

[49] M. Mellody and G. Wakefield, "The time-frequency characteristic of violin vibrato: modal distribution analysis and synthesis," *J. Acoust Soc. Am.*, vol. 107, pp. 598–611, 2000.

[50] M. Mathews, "The digital computer as a musical instrument," *Science*, vol. 142, no. 3592, pp. 553–557, 1963.

[51] J. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *J. Audio Engineering Society*, vol. 21, no. 7, 1973. Reprinted in *Computer Music Journal* 1(2), 1977.

[52] J. Smith, "Viewpoints on the history of digital synthesis," in *Proc. Int. Computer Music Conf. ICMC-91*, (Monreal, Canada), pp. 1–10, 1991.

[53] X. Serra, "Musical sound modeling with sinusoids plus noise," in *Musical Signal Processing* (C. Roads, S. Pope, A. Picialli, and G. D. Poli, eds.), Zwets & Zeitlinger Publishers, 1997.

[54] T. Verma and T. Meng, "Extending spectral modeling synthesis with transient modeling synthesis," *Computer Music J.*, vol. 27, no. 2, pp. 47–59, 2000.

[55] M. Karjalainen and U. Laine, "A model for real-time sound synthesis of guitar on a floating-point signal processor," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal processing*, vol. 5, (Toronto, Canada), pp. 3653–3656, May 1991.

[56] T. Laakso, V. Välimäki, M. Karjalainen, and U. Laine, "Splitting the unit delay – tools for fractional filter design," *IEEE Signal Processing Magazine*, vol. 13, pp. 30–60, 1996.

[57] A. Paladin and D. Rocchesso, "A dispersive resonator in real time on MARS workstation," in *Proc. Int. Computer Music Conf.*, (San Jose, CA), pp. 146–149, 1992.

[58] S. Van Duyne and J. Smith, "A simplified approach to modeling dispersion caused by stiffness in strings and plates," in *Proc. Int. Computer Mus. Conf.*, (Aarhus, Denmark), pp. 407–410, 1994.

[59] D. Rocchesso and F. Scalcon, "Accurate dispersion simulation for piano strings," in *Proceedings of the Nordic Acoustical Meeting NAM 1996*, pp. 407–414, Nordic Acoustical Society, 1996.

[60] M. Karjalainen, V. Välimäki, and T. Tolonen, "Plucked-string models: from the Karplus-Strong algorithm to digital waveguides and beyond," *Computer Music J.*, vol. 22, no. 3, pp. 17–32, 1998.

[61] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy, "Physical modeling of plucked string instruments with application to real-time sound synthesis," *J. Audio Eng. Soc.*, vol. 44, pp. 331–353, 1996.

[62] B. Vercoe, W. G. Gardner, and E. D. Scheirer, "Structured audio: Creation, transmission, and rendering of parametric sound representations," *Proc. IEEE*, vol. 86, no. 5, pp. 922–940, 1998.

[63] E. D. Scheirer and J.-W. Yang, "Synthetic and SNHC audio in MPEG-4," *Signal Processing: Image Communication*, vol. 15, pp. 445–461, 2000.

[64] ISO/IEC, "ISO/IEC JTC1/SC29/WG11 N4031 Multimedia Content Description Interface," 2001.

[65] G. Peeters, S. McAdams, and P. Herrera, "Instrument sound description in the context of MPEG-7." Proc. Int. Computer Music Conf., Berlin, Germany, 2000.

[66] A. Lindsay and W. Kriechbaum, "There's more than one way to hear it: Multiple representations of music in MPEG-7," *Journal of New Music Research*, vol. 28, no. 4, 1999.

[67] T. Tolonen, *Object-based sound source modeling*. PhD thesis, Helsinki University of Technology, 2000.

[68] X. Serra and J. Smith, "Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music J.*, vol. 14, no. 4, pp. 12–24, 1990.

[69] A. Houtsma, "Pitch and timbre: Definition, meaning, and use," *J. New Music Research*, vol. 26, pp. 104–115, 1997.

[70] K. Jensen, *Timbre Models of Musical Sounds*. PhD thesis, Department of Computer Science, University of Copenhagen, 1999.

[71] X. Amatriain, J. Bonada, and X. Serra, "Spectral processing," in *Digital Audio Effects* (U. Zölzer, ed.), Chichester, West Sussex, England: John Wiley & Sons, 2002.

[72] M. Laurson and M. Kuuskankare, "PWSynth: A Lisp-based bridge between computer assisted composition and model based synthesis," in *Proc. Int. Computer Music Conf. ICMC'01*, (Havana, Cuba), pp. 127–130, September 2001.

[73] M. Laurson, V. Välimäki, and C. Erkut, "Production of virtual acoustic guitar music," in *Proc. 22nd AES International Conference*, (Espoo, Finland), pp. 249–255, 2002.

[74] E. Terhardt, G. Stoll, and M. Seewann, "Pitch of complex signals according to virtual pitch theory: Test, examples, and predictions," *Journal of the Acoustical Society of America*, vol. 71, no. 3, pp. 671–678, 1982.

[75] A. J. Houtsma, "Pitch perception," in *Hearing* (B. C. Moore, ed.), ch. 8, Academic press, 1995.

[76] B. C. Moore, "Frequency difference limens for short-duration tones," *J. Acoust. Soc. Am.*, vol. 54, pp. 610–619, 1973.

[77] C. Wier, W. Jesteadt, and D. Green, "Frequency discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.*, vol. 61, no. 1, pp. 178–184, 1977.

[78] E. Terhardt, G. Stoll, and M. Seewann, "Algorithm for extraction of pitch salience from complex tonal signals," *Journal of the Acoustical Society of America*, vol. 71, no. 3, pp. 679–688, 1982.

[79] A. Houtsma, "Noise-induced shifts in the pitch of pure and complex tones," *J. Acoust. Soc. Am.*, vol. 70, pp. 1661–1668, 1981.

[80] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.*, vol. 80, pp. 479–483, 1986.

[81] C. Darwin, V. Ciocca, and G. Sandell, "Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic," *Journal of the Acoustical Society of America*, vol. 95, no. 5, pp. 2631–2636, 1994.

[82] W. M. Hartmann and S. L. Doty, "On the pitches of the components of a complex tone," *Journal of the Acoustical Society of America*, vol. 99, no. 1, pp. 567–578, 1996.

[83] A. de Cheveigné, "Harmonic fusion and pitch shifts of mistuned partials," *Journal of the Acoustical Society of America*, vol. 102, no. 2, pp. 1083–1087, 1997.

[84] J.-Y. Lin and W. M. Hartmann, "The pitch of a mistuned harmonic: Evidence of a template model," *Journal of the Acoustical Society of America*, vol. 103, no. 5, pp. 2608–2617, 1998.

[85] K. von Walliser, "Zusammenhänge zwischen dem schallreiz und der periodentonhöhe," *Acoustica*, vol. 21, pp. 319–329, 1969.

[86] E. Terhardt, "Influence of intensity on the pitch of complex tones," *Acustica*, vol. 33, pp. 344–348, 1975.

[87] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "Thresholds for the detection of inharmonicity in complex tones," *J. Acoust. Soc. Am.*, vol. 77, no. 5, pp. 1861–1867, 1985.

[88] R. Plomp, "Pitch of complex tones," *J. Acoust. Soc. Am.*, vol. 41, pp. 1526–1533, 1967.

[89] J. Schouten, R. Ritsma, and B. L. Cardozo, "Pitch of the residue," *J. Acoust. Soc. Am*, vol. 34, pp. 1418–1424, 1962.

[90] E. de Boer, "On the residue and auditory pitch perception," in *Handbook of Sensory Physiology* (W. Keidel and W. Neff, eds.), berlin: Springer-Verlag, 1976.

[91] P. Vassilakis, “A single model explaining the first and second pitch-shift effects as alternative manifestations of a single phenomenon,” in *Proceedings of the 5th ICMPC*, (Seoul, Korea), pp. 107–114, 1998.

[92] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*. Heidelberg: Springer-Verlag, 1990.

[93] B. Moore and B. Glasberg, “A revision of Zwicker’s loudness model,” *Acustica*, vol. 82, pp. 335–345, 1996.

[94] J. Goldstein, “An optimum processor theory for the central formation of the pitch of complex tones,” *J. Acoust. Soc. Am.*, vol. 54, pp. 1496–1515, 1973.

[95] A. Houtsma and J. Smurzynski, “Pitch identification and discrimination for complex tones with many harmonics,” *J. Acoust. Soc. Am.*, vol. 87, pp. 304–310, 1990.

[96] J. Licklider, “A duplex theory of pitch perception,” *Experientia*, vol. 7, pp. 128–133, 1951.

[97] L. van Noorden, “Two channel pitch perception,” in *Music, Mind and Brain*, London: Plenum, 1982.

[98] A. de Cheveigné, “A pitch perception model,” in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 897–900, 1986.

[99] R. Meddis and M. Hewitt, “Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification,” *Journal of the Acoustical Society of America*, vol. 89, no. 6, pp. 2866–2882, 1991.

[100] P. Cariani and B. Delguette, “Neural correlates of the pitch of complex tones, I: Pitch and pitch salience,” *J. Neurophysiology*, vol. 76, pp. 1698–1716, 1996.

[101] P. Cariani and B. Delguette, “Neural correlates of the pitch of complex tones, II: Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region of pitch,” *J. Neurophysiology*, vol. 76, pp. 1717–1734, 1996.

[102] “Acoustical terminology.” American Standards Association, 1960.

[103] P. Iverson and C. Krumhansl, “Isolating the dynamic attributes of musical timbre,” *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2595–2563, 1993.

[104] J. Kruskall, “Nonmetric multidimensional scaling: A numerical method,” *Psychometrika*, vol. 29, pp. 115–129, 1964.

[105] J. Grey, “Multidimensional perceptual scaling of musical timbres,” *J. Acoust. Soc. Am.*, vol. 61, no. 5, pp. 1270–1277, 1977.

[106] D. Wessel, "Timbre space as musical control structure," *Computer Music J.*, vol. 3, no. 2, pp. 45–52, 1979.

[107] C. Krumhansl, "Why is musical timbre so hard to understand?," in *Structure and perception of electroacoustic sound and music*, Amsterdam: Elsevier, 1989.

[108] S. McAdams, S. Winsberg, G. de Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes," *Psychological Research*, vol. 58, pp. 177–192, 1995.

[109] S. Lakatos, "A common perceptual space for harmonic and percussive timbres," *Perception and Psychophysics*, vol. 62, pp. 1426–1439, 2000.

[110] V. Roussarie, S. McAdams, and A. Chaigne, "Perceptual analysis of synthesized struck bars," *J. Acoust. Soc. Am.*, vol. 103, p. 3007, 1998.

[111] A. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, Massachusetts: MIT Press, 1990.

[112] E. Schorer, "Critical modulation frequency based on detection of AM versus FM tones," *J. Acoust. Soc. Am.*, vol. 79, pp. 1788–1803, 1986.

[113] B. Moore and A. Sek, "Detection of combined frequency and amplitude modulation," *J. Acoust. Soc. Am.*, vol. 92, no. 6, pp. 3119–3131, 1992.

[114] B. Moore and A. Sek, "Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation)," *J. Acoust. Soc. Am.*, vol. 97, no. 4, pp. 2468–2478, 1995.

[115] L. Feth, H. O'Malley, and J. Ramsey, "Pitch of unresolved two-tone complex tones," *J. Acoust. Soc. Am.*, vol. 72, pp. 1403–1412, 1982.

[116] N. Versfeld and A. Houtsma, "Discrimination of changes in the spectral shape of two-tone complexes," *J. Acoust. Soc. Am.*, vol. 98, pp. 807–816, 1995.

[117] E. Zwicker, "Direct comparisons between the sensations produced by frequency modulation and amplitude modulation," *J. Acoust. Soc. Am.*, vol. 34, pp. 1425–1430, 1962.

[118] W. Hartmann and G. Hnath, "Detection of mixed modulation," *Acustica*, vol. 50, pp. 297–312, 1982.

[119] L. Demany and C. Semal, "On the detection of amplitude modulation and frequency modulation at low modulation frequencies," *Acustica*, vol. 61, pp. 243–255, 1986.

[120] B. C. Moore and A. Sek, "Discrimination of frequency glides with superimposed random glides in level," *J. Acoust. Soc. Am.*, vol. 104, no. 1, pp. 411–421, 1998.

[121] J. Madden and K. Fire, "Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency," *J. Acoust. Soc. Am.*, vol. 102, no. 5, pp. 2920–2924, 1997.

[122] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. Los Altos, California: Peninsula Publishing, 1988.

[123] V. Välimäki, M. Laurson, and C. Erkut, "Commuted waveguide synthesis of the clavichord," *Computer Music J.*, vol. 27, no. 1, 2003. In press.

[124] H. Wright, *Acoustics and psychoacoustics of the guitar*. PhD thesis, University of Edinburgh, 1996.

[125] H. Penttinen, M. Karjalainen, T. Paatero, and H. Järveläinen, "New techniques to model reverberant instrument body responses," in *Proc. Int. Computer Music Conf.*, (Havana, Cuba), 2001.

[126] C. Erkut, V. Välimäki, M. Karjalainen, and M. Laurson, "Extraction of physical and expressive parameters fot model-based sound synthesis of the classical guitar," in *Proc. 108th AES Convention*, 2000. Preprint 5114.

[127] W. D'haes and X. Rodet, "Automatic estimation of control parameters: An instance-based learning approach," in *Proc. Int. Computer Music Conf.*, (Havana, Cuba), pp. 199–202, 2001.

[128] A. Nackaerts, B. De Moor, and R. Lauwereins, "Parameter estimation for dual-polarization plucked string models," in *Proc. Int. Computer Music Conf.*, (Havana, Cuba), pp. 203–206, 2001.

[129] J. Riionheimo and V. Välimäki, "Parameter estimation of a plucked string synthesis model with a genetic algorithm," in *Proc. Int. Computer Music Conf. ICMC2002*, (Gothenburg, Sweden), pp. 283–286, 2002.

[130] A. Chaigne, "Numerical simulations of stringed imstruments – today's situation and trends for the future," *Catgut Acoustical Society Journal*, vol. 4, no. 5, pp. 12–20, 2002.

[131] S. Canazza, G. De Poli, A. Rodá, A. Vidolin, and P. Zanon, "Kinematics-energy space for expressive interaction in music performance," in *Proc. Workshop on Current Research Directions in Computer Music*, (Barcelona), November 2001.