The Timbre Model

Kristoffer Jensen

Music Informatics Laboratory
Department of Datalogy, University of Copenhagen
Universitetsparken 1, 2100 Copenhagen Ø, Denmark
krist@diku.dk, http://www.diku.dk/~krist

Abstract

This paper presents the timbre model, a signal model which has been built to better understand the relationship between the perception of timbre and the musical sounds most commonly associated with timbre. In addition, an extension to the timbre model incorporating expressions is introduced. The presented work therefore has relation to a large field of science, including auditory perception, signal processing, physical models and the acoustics of musical instruments, music expression, and other computer music research. The timbre model is based on a sinusoidal model, and it consists of a spectral envelope, frequencies, a temporal envelope and different irregularity parameters. The paper is divided into four parts: an overview of the research done on the perception of timbre, an overview of the signal processing aspects dealing with sinusoidal modeling, the timbre model, and an introduction of some expressive extensions to the timbre model.

1 Introduction

The sound of the musical instrument can be qualified by the timbre (or the identity) of the sound and the expressions caused by gestures. Expressions associated with musical instruments are well defined by common musical terms, such as note, intensity, tempo or style. Timbre seems to be a multi-dimensional quality. Generally, timbre is separated from the expression attributes pitch, intensity, and length of a sound. Furthermore, research has shown that timbre consists of the spectral envelope, an amplitude envelope function, which can be attack, decay, etc., the irregularity of the amplitude of the partials, and noise.

The additive parameters constitute a good analysis/synthesis model of voiced sounds, but it has a very large parameter set, which is not always easily manipulated. This paper presents an approach to modeling the additive parameters in an easily controllable, intuitive model, whose parameters are closely related to the timbre dimensions as proposed in timbre research

The timbre model models each partial in a few pertinent parameters: the spectral envelope, the mean frequencies, the amplitude envelope, and the noise, or irregularity parameters. Furthermore, the rate and extend of the vibrato and tremolo are modeled.

The timbre model has a fixed parameter size, dependent only on the number of partials, and most of the parameters of the model have an intuitive perceptive quality, due to their relation with timbre perception. Furthermore, by interpolating between parameter sets obtained from different sounds, morphing is easily implemented.

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

This paper starts with an overview of the perceptual research, which forms the basis for the model, then an overview of the additive analysis is given, and the timbre model is detailed. Next a novel set of expression additions to the timbre model is presented, other applications to the model are outlined and a conclusion is given.

2 Perceptual Research

The timbre model is derived from conclusions extracted from the auditory perception research. Several methodologies have been used in this research, and even though the results are given in many formats, the conclusions are generally the same. The research suffers from a lack of comparable sound material, and in particular, the lack of noisy, non-harmonic or percussive sounds in the experiments. In addition, different pitches are generally not used.

This is an overview of the research on which the timbre model is based. For a larger overview of timbre research, see for instance [38], [65].

In a larger scope, [7] presents some aspects of timbre used in composition. Timbre research is of course related to auditory perception [102] and psycho-acoustics [105] research. An initial study of the Just Noticeable Difference (JND) of many of the timbre attributes presented here can be found in [48].

The mpeg community, which defined the popular mp3 compression standard, are currently defining mpeg 7 [62], which defines the content of sound and music. The outcome of this standardization process may be very useful in finding common terms for objectively describing music and music sounds. It differentiates between noise and harmonic sounds, substituting spectrum measures in the first case with harmonic measures in the second case.

2.1 Timbre Definition

Timbre is defined in ASA [2] as that quality which distinguishes two sounds with the same pitch, loudness and duration. This definition defines what timbre is not, not what timbre is.
Timbre is generally assumed to be multidimensional, where some of the dimensions has to do with the spectral envelope, the amplitude envelope, etc. The difficulty of timbre identity research is often increased by the fact that many timbre parameters are more similar for different instrument sounds with the same pitch, than for sounds from the same instrument with different pitch. For instance, many timbre parameters of a high pitched piano sound are closer to the parameters of a high-pitched flute sound than to a low-pitched piano sound. Nevertheless, human perception or cognition generally identifies the instrument correctly. Unfortunately, not much research has dealt with timbre perception for different pitches. Greg Sandell has made a list [84] of different peoples definition of the word timbre.

2.2 Verbal Attributes

Timbre is best defined in the human community outside the scientific sphere by its verbal attributes (historically, up to and including today, by the name of the instrument that has produced the sound). von Bismarck [98] had subjects rate speech, musical sounds and artificial sounds on 30 verbal attributes. He then did a multidimensional scaling on the result, and found 4 axes, the first associated with the verbal attribute pair dull-sharp, the second compact-scattered, the third full-empty and the fourth colorful-colorless. The dull-sharp axis was further found to be determined by the frequency position of the overall energy concentration of the spectrum. The compact-scattered axis was determined by the tone/noise character of the sound. The other two axes were not attributed to any specific quality.

2.3 Dissimilarity Tests

The dissimilarity test is a common method of finding proximity in the timbre of different musical instruments. Asking subjects to judge the dissimilarity of a number of sounds and analyzing the results is the essence of the dissimilarity tests. A multidimensional scaling is used on the dissimilarity scores, and the resulting dimensions are analyzed to find the relevant timbre quality.
Grey [33] found the most important timbre dimension to be the spectral envelope. Furthermore, the attack-decay behavior and synchronicity were found important, as were the spectral fluctuation in time and the presence or not of high frequency energy preceding the attack.
Iverson & Krumhansl [44] tried to isolate the effect of the attack from the steady state effect. The surprising conclusion was that the attack contained all the important features, such as the spectral envelope, but also that the same characteristics were present in the steady state. The resulting timbre space was similar no matter if the full tones, the attack only or the remainders only were examined.
Later studies Krimphoff *et al.* [54] refined the analysis, and found the most important timbre dimensions to be brightness, attack time, and the spectral fine structure.
Grey & Gordon [35], Iverson & Krumhansl [44] and Krimphoff *et al.* [54] compared the subject ratings with calculated attributes from the spectrum. Grey & Gordon [35] found that the centroid of the bark [89] domain spectral envelope correlated with the first axis of the analysis. Iverson & Krumhansl [44] also found that the centroid of the spectral envelope, here calculated in the linear frequency domain (brightness), correlated with the first dimension. Krimphoff *et al.* [54] also found the brightness to correlate well with the most important dimension of the timbre. In addition, they found the log of the rise time (attack time) to correlate with the second dimension of the timbre, and the irregularity of the spectral envelope to correlate with the third dimension of the timbre. McAdams *et al.* [65] further refined this hypothesis, substituting spectral irregularity with spectral flux.
The dissimilarity tests performed so far do not indicate any noise perception. Grey [33] introduced the high frequency noise preceding the attack as an important attribute, but it was later discarded in Iverson & Krumhansl [44]. The lack of indications of noise discrimination might be explained by the fact that no noisy sounds were included in the test sounds. McAdams *et al.* [65] promises a study with a larger variety of test sounds, including noisy sounds. It can also be explained by the fact that the most common analysis methods doesn't permit the analysis of noise, which then cannot be correlated with the ratings.

2.4 Auditory Stream Segregation

An interesting way of examining the qualities of timbre that can be related to its perception is the auditory stream segregation [12].
The auditory stream segregation is referring to the tendency to group together (i.e. relate them to the same source) sounds with components falling in similar frequency ranges. This phenomenon is called fusion or coherence. The opposite phenomenon when the sounds are separated into several groups, as coming from different sound sources, is called fission or segregation. For intermediate frequency separations between successive tones in a rapid sequence the percept may be ambiguous.
Experimenters have taken advantage of auditory stream segregation to identify timbre qualities related to timbre perception. Using a three tone repeated sequence, by means of adjusting spectral characteristics of the middle tone they tried to investigate how the fusion/fission boundary is affected in relation to the value it assumes for the monotimbral (no change in the timbre of the middle tone) case.
Singh and Bregman [91] presented monotimbral and bitimbral sequences of complex sounds to listeners

where for the bitimbral case there were changes in the attack time and the number of harmonics in the middle sound. The results indicate that the effect of the change was highly significant for both the fission and the fusion boundary fundamental frequency value. The boundaries assumed lower values than for the monotimbral case in all changes and the order of the impact on the results was higher for changes in the number of partials than for the envelope changes. This method seems promising when searching for an absolute value for timbre changes, since any timbre change provoking fusion/fission can be related to the corresponding pitch change when there is no timbre change.

2.5 Discrimination

Several studies asked subjects the rate the differences between original (unmodified) sounds and modified sounds, in order to evaluate the perceptual importance of the modification. One recent such study is McAdams *et al.* [64], in which the additive parameters of seven sounds are modified in different ways, and the importance of the modifications are asserted. The sounds, clarinet, flute, oboe, trumpet, violin, harpchichord and marimbe were normalized to E4-flat (311.1 Hz), 2 secs and equal subjective loudness. The participants were asked to discriminate between the original (the normalized E4-flat, 2secs) and modified sounds, and some of the results of this research are:

The effect of musical training is weak but the effect of instrument is strong. Therefore each instrument is analyzed individually whereas the participants are grouped.

The most important results from the effect of simplification are that most simplifications are perceptible and that accumulated simplifications are equivalent to the most potent constituent (Most salient feature dominate in combined simplifications).

The order of the simplifications are: spectral envelope smoothing, spectral flux (amplitude envelope coherence) (very good discrimination), forced harmonic frequency variations, frequency variations smoothings, frequency micro-variations and amplitude micro-variations (moderate to poor discriminations).

In conclusion, several research methods have been used to determine the dimensions of timbre. Although no clear consensus has emerged, the most common dimensions seem to be spectral envelope, temporal envelope and irregularities.

3 Additive Analysis/Synthesis

The additive model has been chosen as the basis for the timbre model for the known analysis/synthesis qualities and the perceptually meaningful parameters of this model. Many analysis/synthesis systems using the additive model exist today, including sndan [8], SMS [90], the loris program [25] and the additive program [81].

3.1 Choice of model

In order to perform analysis by synthesis, the timbre model must be based on an analysis/synthesis model. The choice of underlying model is the additive (sinusoidal) model, for its well-understood parameters (time, amplitude and frequency) and for its proven analysis methods. Other methods investigated include the physical model [45], the granular synthesis [96], the wavelet analysis/synthesis [55], the atomic decomposition [14], [36] and the modal distribution analysis [72].

In the additive analysis, [90] added a stochastic part to the sound, making the model non-homogenous. Other non-homogenous additions include the transients [97].

3.2 Additive Model

The additive analysis consists in associating a number of sinusoidal with a sound, and estimating the time-varying amplitudes ($a_k(t)$) and frequencies ($f_k(t)$) of the N sinusoidals (partials) from the sound. The sound can then be resynthesized, with a high degree of realism, by summing the sinusoidals,

$$s(t) = \sum_{k=1}^{N} a_k(t) \ \sin(2\pi \int_{\tau=0}^{t} f_k(\tau) d\tau) . \qquad (1)$$

In order to have a good resynthesis quality, the absolute phases are generally necessary [3].

3.3 Additive Analysis

Several methods exist for determining the time-varying amplitudes and frequencies of the partials. Already in the last century, musical instrument tones were divided into their Fourier series [40], [73]. Early techniques for the time-varying analysis of the additive parameters are presented by Matthews *et al.* [63] and Freedman [28]. Other more recent techniques for the additive analysis of musical signals are the proven heterodyne filtering [34], the much-used FFT-based analysis [66], and the linear time/frequency analysis [37]. The time and frequency reassignment method [4] has recently gained popularity [11], [23]. Ding and Qian [19] has presented an interesting method, fitting a waveform by minimizing the energy of the residual, improved and dubbed adaptive analysis by Röbel [78]. A great deal of research has been put into understanding and improving [99], [17] the windows [39] of the short term fourier transfer [1]. Finally, Marchand [59] used signal derivatives to estimate the amplitudes and frequencies.

3.4 Analysis/synthesis evaluation

Not many objective listening tests (outside the compression community) have been performed in the music community to evaluate analysis/synthesis methods. [94] evaluated the spectral/time envelope model with listening tests. [34] compared analysis/synthesis and different data-reductions, and [84] evaluated the PCA-based data reduction with listening tests.

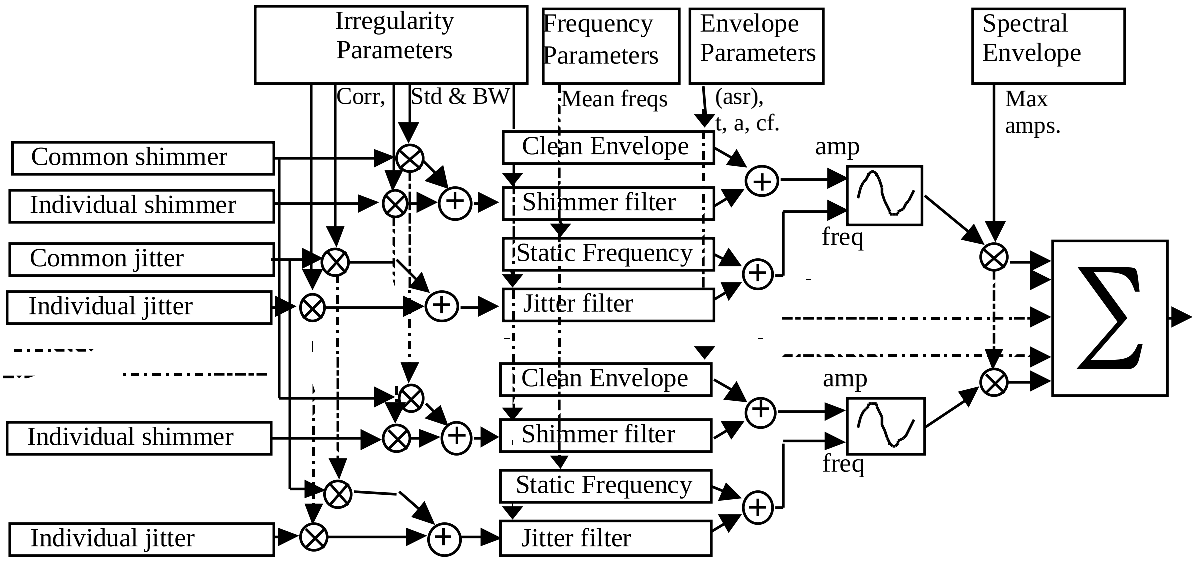

Figure 1. Timbre model diagram. The model consists of a number of sinusoidals, with amplitude as a sum of the clean envelope and the shimmer multiplied with the spectral envelope, and the frequency as a sum of the static frequency and the jitter. The shimmer and jitter are a sum of common (correlated) and individual gaussian white noise filtered and scaled by the standard deviation.

Recently, however, the additive analysis/synthesis has been tested in several well-controlled listening test experiments. The listening tests performed in connection with this work have been inspired by the listening tests performed for the evaluation of speech and music compression. The method used is called double blind triple stimulus with hidden reference [74]. [52] found the linear time-frequency additive analysis/synthesis to have a sound quality between imperceptible and perceptibel, but not annoying using short musical sounds. [13] found comparable results using several additive analysis systems and longer musical sequences. In addition [3] found that including the phase gave significantly better results. Other evaluation methods include estimating the time resolution, [52] found that the LTF analysis method [37] has a time resolution that is twice as good as the FFT-based analysis, and error estimations. [11] found the mean error in the amplitude and frequency estimation to be significantly better for the frequency reassignment method, than the peak interpolation, or phase difference methods.

In this work, the FFT-based analysis is used. The additive parameters are saved for each half-period of the fundamental, and only quasi-harmonic partials are estimated. The window used is the kaiser window, and the block size is 2.8 periods of the fundamental.

4 The Timbre Model

The timbre model is inspired by the perceptual research, but it has been derived from the analysis of musical sounds using the analysis by synthesis method [77] and by literature studies into related research. By these methods, the model has been defined to consist of a spectral envelope, associated with brightness and resonances of the sounds, a frequency envelope, associated with pitch and inharmoncity. It also consists of an amplitude envelope consisting of five segments, start, attack, sustain, release and end, each with individual start and end relative amplitude and time, and curve form, and the amplitude and frequency irregularity (shimmer and jitter). The shimmer and jitter are modeled as a low-pass filtered gaussian with a given standard deviation and bandwidth. The amplitude envelope is associated with important timbre attributes, such as attack time, and sustained/percussive quality, and the irregularity is associated with both additive noise, but also slow irregularities, giving life to the sounds.

Other methods of modeling the additive parameters include the Group Additive Synthesis [53], [22], [15], where similar partials are grouped together to improve efficiency. [95] use envelope time points to morph between different musical sounds. Marchand has proposed the Structured Additive Synthesis [58].

Schaeffer proposes a classification of attack genres (among others things) in his monumental work [86].

The base of the timbre model is the additive parameters, the amplitude of which is controlled by the spectral envelope, amplitude envelope and shimmer parameters, and the frequency of which is controlled by the mean frequency and the jitter parameters.

The timbre model diagram can be seen in figure 1. It consists of a number of sinusoidals (partials), whose amplitude is the sum of a clean envelope (attack-sustain/decay-release) and irregularity (shimmer) multiplied with the spectral envelope value, and whose frequency is the sum of a static value and irregularity (jitter). The timbre model parameters are (from right to left): Max amplitudes, envelope model times, amplitudes and curve form coefficients, mean frequencies, irregularity correlation, standard deviation and bandwidth.

4.1 The Spectral Envelope

The spectral envelope is very important for the perceived effect of the sound; indeed, the spectral envelope alone is often enough to distinguish or recognize a sound. This is especially true for the recognition of vowels, which are entirely defined by the spectral envelope. As was seen earlier, the spectral envelope is indeed one of the most important timbre dimensions. Nevertheless, the spectral envelope alone is not enough to recreate any sound with realism. Many methods exist to model the spectral envelope, including the linear predictive coding (lpc), cepstrum, etc. [88]. Back in 1966 [94] synthesized wind instruments with a combination of spectral and temporal envelopes. [80] use spectral envelopes as a filter with different source models, including the additive model.

[70] introduced the discrete summation formulas in sound synthesis, which are here called the brightness creation function [52]. There exists an easy calculation and recreation of brightness with these formulas [52].

The spectral envelope is defined in this work as the maximum amplitude of each partial, denoted $\hat{a}_k$. As it is difficult to estimate the spectral envelope outside the discrete frequency points in voiced sounds, the spectral envelope model using the partial amplitudes is judged the most reliable.

4.2 Frequencies

The static frequencies are modeled as the weighted mean of the frequency for the sustain part of each partial, denoted $\hat{f}_k$.

Most sustained instruments are supposed to be perfectly harmonic, i.e. $\hat{f}_k = k\hat{f}_0$. The frequencies are therefore best visualized divided by the harmonic partial index. The piano, in contrast, has inharmonic partial frequencies due to the stiffness of the strings [26]. Therefore, the piano partial frequencies are slightly higher than the harmonic frequencies.

4.3 Amplitude Envelopes

The envelope of each partial is modeled in five segments, start and end segments, supposedly silent, and attack, sustain and release segments. Thus, there are 6 amplitude/time split points, where the first is (0,0) and the last amplitude also is zero, since all partials are supposed to start and end in silence. The amplitudes are saved as a percentage of the maximum of the amplitude (the spectral envelope), and the times are saved in msec. Furthermore, the curve form of each segment is modeled by a curve, which has an appropriate exponential or logarithmic form. The resulting concatenated clean amplitude envelopes are denoted $\tilde{a}_k(t)$. The formula for one segment can be seen in equation (13).

The envelope split-points and curve form coefficients are found by a method inspired by the scale-space community in image processing [56], in which the split-points are found on the very smoothed time-derivative envelopes, and then followed gradually to the unsmoothed case. The smoothing is performed by convoluting the envelope with a gaussian,

$$env_\sigma(t) = a_k(t) * g_\sigma(t), \quad g_\sigma(t) = \frac{1}{2\pi\sigma} e^{-\frac{t^2}{2\sigma^2}}. \quad (2)$$

The derivative of the amplitude has been shown to be important in the perception of the attack [32]. The middle of the attack and release are now found by finding the maximum and minimum of the time derivative of the smoothed envelope,

$$\frac{\max}{\min} L_{t,\sigma}(t), \qquad L_{t,\sigma}(t) = \frac{\partial}{\partial t} env_\sigma(t) \qquad (3)$$

The start and end of the attack and release are found by following $L_{t,\sigma}$ forwards and backwards (in time) until it is close to zero (about one tenth of the maximum derivative for the attack and end of release, and the double for the start of release). This method generally succeeds in finding the proper release time for the decay-release piano sound. A further development of the envelope analysis and model can be found in [52], [47].

4.4 Irregularities

Although the clean recreated envelopes have the general shape of the original envelope, there is a great deal of irregularity left, which is not modeled in the clean envelopes. The same holds true for the frequencies. The irregularities are divided into periodicity and non-periodic noises. The noise on the amplitude envelope is called shimmer, and the noise on the frequency is called jitter. Shimmer and jitter are modeled for the attack, sustain and release segments. It is supposed to have a Gaussian distribution; the amplitude of the noise is then characterized by the standard deviation. The frequency magnitude of the noise is assumed low-pass and modeled with the 3dB bandwidth, and the correlation between the shimmer and jitter of each partial and the fundamental is also modeled.

Other noise models of musical sounds include the residual noise in the FFT [90], [69] and the random point process model of music [76] or speech noise [75].

Models of noise on sinusoidals include the narrow band basis functions (NBBF) in speech models [61]. In music analysis, [24] introduced the bandwidth enhanced sinusoidal modeling. Both models model only jitter, not shimmer.

Other analysis of the noise, and irregularity of the music sounds include the analysis of aperiodicity [67], [87], and the analysis of higher order statistics [20], [21].

The shimmer and jitter are calculated normalized with the clean amplitudes and the mean frequencies respectively,

$$shimmer_k = \frac{a_k(t) - \tilde{a}_k(t)}{\hat{a}_k} \qquad (4)$$

$$jitter_k = \frac{f_k - \hat{f}_k}{\hat{f}_k} \quad (5)$$

The jitter and shimmer is then assumed to be stochastic, with a gaussian distribution, and modeled by its standard deviation and bandwidth (and later recreated using a single-tap recursive filter).

4.5 Vibrato and Tremolo

Although assumed to be part of the expression of the sound, and therefore not necessary in the timbre model, some periodicity is nevertheless often found in the time-varying amplitudes and frequencies of the partials. This periodicity is modeled by its rate, strength, and phase. The frequency periodicity is here called vibrato, and the amplitude periodicity is called tremolo. There are individual vibrato and tremolo parameters for each partial.

The vibrato and tremolo parameters are found by searching for the max in the absolute value of the FFT of the time-varying frequency (subtracted by the mean frequency of each partial) or amplitude (subtracted by the clean amplitude curve). This provides an estimate of the strength, rate and phase of the periodic information in the signal, if there is any. [41] uses a comparable method for the vibrato estimation.

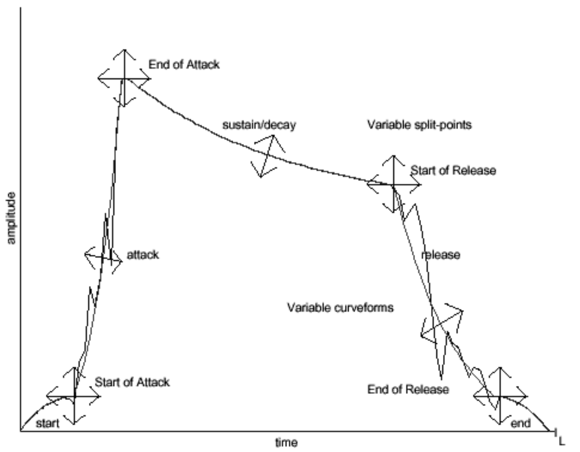

Figure 2. Total amplitude envelope for one partial The five segments (start, attack, sustain, release and end) have individual split-points and curve-forms. The attack has high bandwidth shimmer, and the release has low bandwidth shimmer.

With the addition of irregularities, the timbre model is finished. An example of the resulting amplitude envelope for one partial can be seen in figure 3. The envelope is slightly decaying, and it has high bandwidth shimmer at the attack, and low bandwidth shimmer at the release.

4.6 Visualization

The timbre model parameters have now been presented, and an overview of the parameter estimation methods has been given. A more throughout presentation of the timbre model is given in [52]. This section presents a proposed visualization of the timbre attributes. Many of them are best plotted logarithmically, and all the attributes are plotted with fixed axes, to facilitate comparisons between sounds.

There are 12 different timbre attributes, some of which have values for more than one segment (attack, decay, sustain, release or end). In order to have an easy overview, all the timbre attributes are collected in one plot. All the start and end (which are assumed to be silent) parameters are omitted. In total, there are 12 attributes, which can be plotted in one figure in 6 rows and 2 columns. The left column has from the top to the bottom the spectral envelope, the normalized frequencies divided by the partial index and fundamental, and envelope timing (attack and release), the envelope percents, the envelope curve forms, and the vibrato and tremolo rate. The right column has from the top to the bottom the shimmer standard deviation, the jitter standard deviation, the shimmer filter bandwidth, the jitter filter bandwidth, the shimmer and jitter correlation and the tremolo/vibrato strength.

5 Expressive Additions to the Timbre Model

The timbre model has proven its value in a number of applications, including the analysis of different expression styles [52]. This analysis will here serve as the basis for the inclusion of a number of parameters, which govern the behavior of the timbre model attributes when used in a synthesis context. The expressions are treated in an analysis/synthesis context, and adapted for real-time synthesis [60], where possible.

The expressive modes introduced to the timbre model include variants, pitch, intensity, vibrato and tremolo, and other expressions, such as legato/staccato.

The attempt to find expressive parameters that can be included in the timbre model is an initial study. Both additional literature studies and field test using the timbre engine [60] must be performed in order to assert the validity of the expression model. Of course, the music performance studies [31] gives a lot of information, which can also be gathered from research dealing with the acoustics [5], [9] or the physics [27] of musical instruments. In addition, appropriate gestures must be associated with the expression parameters [51], [100], [101] [43]. The vibrato problem [18] is an interesting topic, stating, for instance, whether to add cycles (vibrato) or stretch (glissando) a time-scaled continuous expression parameter. This is not a problem in the timbre model, because of the division into clean envelopes, where only the sustain part is scaled, and periodic and non-periodic irregularities, which are not scaled.

5.1 Variants

In this work, the sound of a musical instrument is divided into an identity part, and an expression part, where the identity is the neutral sound, or what is

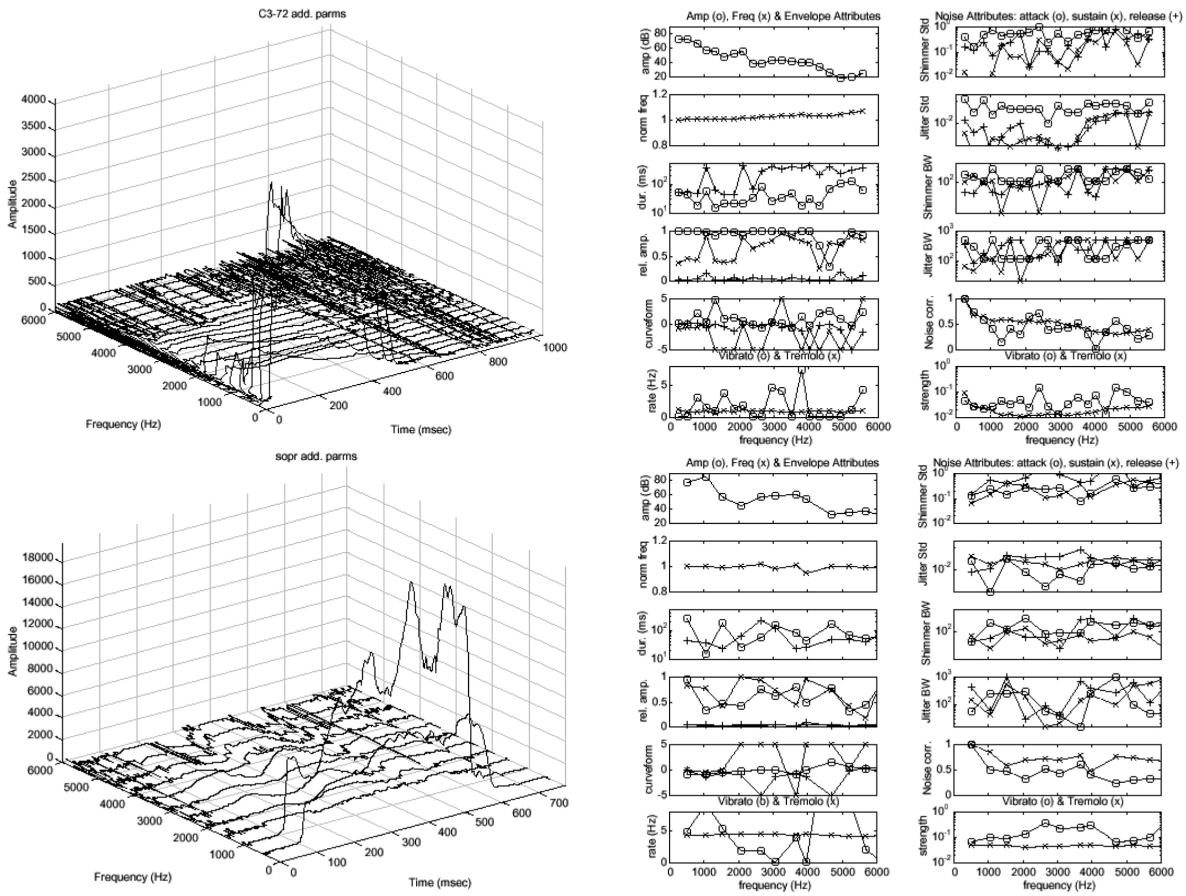

Figure 3. Additive parameters (left) and corresponding timbre attributes (right), Piano (top) and Soprano (bottom). Since all the axes in the timbre attribute plots are fixed, it is easy to compare the sounds. The amplitude values are denoted (o) and the frequency values (x). Attack is (o), sustain (x) and release (+). The piano and soprano have resonances in the spectral envelope, the piano is slightly inharmonic, and it has a faster attack. The piano is definitely percussive, since the attack relative amplitude is higher than the release. The soprano has a definite vibrato at around 4 Hz. Both the piano and the soprano have tremolo on some partials. The piano has much band-pass jitter in the attack. The jitter is more correlated than the shimmer for both sounds.

common in all the expression styles of the instrument, and the expression is the change/additions to the identity part introduced by the performer. The expression can be seen as the acoustic outcome of the gesture manipulation of the musical instrument. This division, however, it not simple in an analysis/synthesis situation, since the sound must be played to exist, and it thereby, by definition, always contains an expression part. One attempt at finding the identity is by introducing the notion of variants, which is assumed to be the equivalent of the sounds coming from several executions of the same expressive style.

The variants are calculated by introduced an ideal curve to the different timbre attributes. The deviation from this ideal curve is then assumed to be stochastic, with a given distribution, and for each new execution, a new instance of the deviation is created, giving the sound a clearly altered timbre. Some of the timbre attributes have ideal curves corresponding to some perceptual or physical reality, such as the brightness creation function [52] for the spectral envelope,

$$a_k = a_0 \left(\frac{B}{B-1} \right)^{-k} \tag{6}$$

where a_k is the amplitude of the partial k, a_0 is the fundamental amplitude, and B is the estimated brightness [6], see equation (9), and the equation for the ideal stiff string for the frequencies [26],

$$f_k = k f_0 \sqrt{1 + \beta k^2} \tag{7}$$

where β is the inharmonicity coefficient. Studies of the discrimination of inharmonicity can be found in, for instance, [46] and [79].

Most timbre attributes, however, are fitted with a simple exponential curve,

$$c_k = v_0 \ e^{v_1 k} \tag{8}$$

where v_0 is the fundamental value and v_1 is the exponential coefficient. This curve can model both almost linear curves with small v_1, but also exponential behaviors.

The parameters of the curves are found by minimizing the lms error using the Levenberg-

Marquardt algorithm [71], except for the spectral envelope curve, which is created from the estimated brightness [6],

$$B = (\sum_{k=1}^{N} k a_k) / \sum_{k=1}^{N} a_k \tag{9}$$

The deviations from the ideal timbre attributes parameters are now calculated as,

$$d_k = c_k - \hat{c}_k \tag{10}$$

where c_k are the estimated timbre attribute parameters (amplitude, frequency, or other parameter), and $\hat{c}_k$ are the ideal parameters and d_k is the deviation (assumed to be white gaussian noise).

The deviation d_k between the clean exponential curve and the estimated attributes is assumed to be related to the execution of the sound, and the error can, if modeled properly, introduce new executions of the same sound, i.e. of the same instrument, player and style, in the same environment.

Although the clean curves generally fit well with the estimated parameters, there can be discrepancies caused by bad parameter estimation, correlated deviations between attributes, or inadequate modeling. These discrepancies generally do not make artifacts in the original timbre attribute generated sounds, but they sometimes make the variant sounds too different. One way of minimizing this phenomenon is by using weighted curve-fits, which does remove some of the discrepancies. However, since the heavily different variants may be desired, the variants influence is scaled,

$$\tilde{c}_k = \hat{c}_k + v \ \hat{d}_k + (1 - v) \ d_k \tag{11}$$

where a variant scaling (v) of zero gives the original timbre attributes, and a scaling of one gives entirely new timbre attribute deviations ($\hat{d}_k$) for each execution. The total deviations can additionally be weighted [60], permitting more, or less, deviations from the identity of the sound.

5.2 Pitch

The modeling of the pitch evolution of the timbre attributes is important when executions for different notes are not available. This is the case, for instance, when creating new, or altered, timbre model parameters sets. Obviously, it's impossible to make pitch rules that encompasses all possible musical instruments, so instead, a simple interpolation scheme has been devised, which assures the proper modification of at least the important spectral envelope. In this scheme, all the timbre attributes are assumed to have the partial frequencies in the x axis, and the timbre attributes for the new pitch is found by interpolating between the neighboring values for each partial frequency. The values outside the original frequencies are found by using the extreme values. Although this scheme is rather simple, it has the double advantage of handling the most important timbre attribute correctly, and assuring continuous variations in the resulting sound, as the pitch is altered. In addition, the new timbre attribute values should be interpolated after each pitch change, thereby allowing for subtle effects, caused by for instance resonances. When adding vibrato, the sound would have a continuously varying timbre, thereby adding more life to the execution.

The scheme does not handle sound level, however. In the work on generative rules for music performance, [29] suggest a rule, which raises the sound level 3 dB/octave. An initial study have shown that this effect is present in many musical instruments, although ranging from below 3 (violin) to around 3 (piano, clarinet) to 10 (flute), and 15 dB/octave (soprano). This effect is therefore parameterized (N dB/octave) and included into the expressive model.

5.3 Intensity

The intensity, or velocity, is another important expression attribute. In general, when increasing the velocity, blowing force, or bow velocity, etc., two things happen, the sound emitted gets louder, and it gets brighter. [10] showed that the increase in amplitude and brightness is asymptotic, i.e. the value changes less as velocity grows. In addition, it was shown that the change of brightness in the piano when changing the velocity of the hammer is governed by a straight line in the Hz/dB domain. Therefore, this is the model chosen for the intensity. The amplitude and spectral tilts (slope of the straight line) have an exponential form,

$$val = (v_M - v_m \ e^{-\beta \ v}), \tag{12}$$

where val can be either amplitude or spectral tilt [10], and v_M and v_m defines the maximum and minimum values, β the sensibility and v is the velocity.

The values of v_M, v_m and β can be determined from the instrument, if enough velocity executions are available [10], or it can be user defined. In particular, the upper spectral tilt slope is a useful expression parameter, since it defines how bright the sound becomes, when increasing the velocity to a maximum. This model is also consistent with the analysis of piano executions performed in [52], which showed that the change of velocity (of the piano hammer) only affected the spectral envelope, except for an as yet unexplained change in the shimmer and jitter correlations.

5.4 Duration

The duration is of course also a very important expression parameter. Since the timbre model encompasses both percussive and sustained sounds, a general strategy for modifying the length of the sound is necessary. This strategy could be found by following the clean curve of the sustain/decay part of each partial, with the given curve form,

$$\tilde{a}_k(t) = a_0 + (a_T - a_0) \frac{c^{t/T} - 1}{c - 1}, \tag{13}$$

where c is the curve form coefficient, T is the segment duration, and a_0 and a_T are the start and end values. Unfortunately, the curve form is sometimes fitted to part of the release segment, or sometimes only part of the sustain/decay segment is used, therefore the curve form is error prone. Instead, the decay is assumed to be logarithmic, and modeled as a straight line in the dB domain,

$$\hat{a}^{dB}(t) = \hat{a}_0^{dB} + bt, \tag{14}$$

where $\hat{a}_0$ and b are found by fitting to the known split-point amplitude/time values. If the execution is lasting (played a long time), and the second split-point has a higher amplitude than the first one (b>0), then clipping will eventually occur. The perceptual effect of this has been judged to be interesting enough to not prevent this happening. Instead, the amplitude of each partial is limited to a value at a user-defined percentage above the maximum value of the partial. Obviously, if this limit is set to 100%, then no crescendo effect is possible.

[29] tentatively suggests a modification to the attack and decay in durational contrasts. This is an interesting inclusion, but this effect has not been found [47] in relation to this work (these durational contrasts were not part of the sound material under test), and it is not included in the model.

5.5 Vibrato and Tremolo

The vibrato and tremolo are important expression parameters with specific values defined by the musical context in which the execution is produced. Therefore, a generative rule for the addition of vibrato and tremolo is not desirable in this work. Some vibrato or tremolo effects are, however, part of the identity of the sound, and these effects should not be user-controlled, but inherent in the sound. In particular, this is the case for the amplitude modulation caused by the beating of different modes or strings in, for instance, the piano [103].

The vibrato and tremolo are generally defined by three parameters, the rate (speed), the strength and the initial delay. [31] reviews several vibrato studies, and reports the vibrato rate of professional singers to be between 5.5 to 8 Hz, with a strength between one-tenth of a full tone to one full tone, averaging 0.5 to 0.6 of a whole-tone step.

[82] models the vibrato with the sum of a number of sinusoidals with time-varying amplitudes and phases. The phase of the vibrato/tremolo is necessary, if the resulting sound should be perceptually identical to the original one. Care must be taken to model the resonances correctly when adding vibrato [68]. In addition, the perceived pitch of tones with vibrato is also an important research field [16].

In order to assert whether a sound contains a vibrato or tremolo expression, or whether it contains periodic vibrations in its identity, two things can be examined. First, if the partials are heavily correlated, secondly, if the rate and strength values are correlated, then the chances of it containing expressive vibrato/tremolo is great. If neither of the two cases occur, periodicity is assumed to be part of the identity of the sound, and not controlled by the performer. If expression periodicity is found, it is removed from the sound, and only added back, if and when the performer is signifying it.

5.6 Other expressions

The expressions can be any kind of acoustic change resulting from manipulation of the music instrument.

The other expression parameters used in classical music include mainly styles (legato/staccato, for instance) and tempi. Since some of these expressions are controlled continuously by the performer, they are not easily integrated into the timbre model. In particular, no good gesture capture device has been available to perform tests. In addition, not much timbre attribute changes have been found when analyzing executions of different styles [52]. Therefore, the conclusion must be that the styles are mainly a matter of duration, which is easily controlled in this model.

Another important expression is the transition [93]. Since the transition is the time-varying amplitude, fundamental frequency, and timbre, it should not be too difficult to create timbre attribute sets with appropriate values for different transition.

Finally, another possible expression is the generic timbre navigation [104], [83]. In this case, the timbre is manipulated using sensors and various mapping strategies [51], [101].

6 Other Applications

The timbre model has been successfully used in many applications, in particular in classification and sound manipulation, including morphing.

6.1 Classification

Classification is an important topic today, for instance with the growing number of on-line sound samples.

The timbre model was used as the basis for sound classification [52], in which a subset of the timbre attributes (16 attributes from each sound) was used to classify 150 sounds in 5 instrument classes with no errors. Binary tree classification, using approximately the same data set, was presented in [49], giving much information about the importance of the timbre attributes in the classification.

A real-time learning classification scheme was presented in [30]. For a recent overview of musical instrument classification methods, see [42].

6.2 Sound manipulation

Morphing and sound manipulation is another application in which the timbre model has shown its value. Since the timbre model parameter set has a fixed size (except for the number of partials), it is easy to morph between the sounds by simply interpolating different timbre sets. The interpolated timbre model parameters can also be used to

manipulate the additive parameters [52]. A similar strategy, but with an unequal number of features was presented in [95]. [83] interpolates the additive parameters directly, and [88] uses the spectral envelope to modify sounds. Other morphing strategies include the Karaoke impersonator [57]. [92] uses results from auditory scene analysis to warp and interpolate between two sounds.
Finally, a simplified timbre model [52] was presented in [50] for use in prototyping of musical sounds.

7 Conclusion

The paper gives an overview of the perceptual research and analysis/synthesis of the additive parameters. It presents the timbre model, a signal model, which models most musical sounds with a high fidelity. The timbre model is based on perceptual research and analysis by synthesis. It consists of a spectral envelope, frequencies, a five-segment amplitude envelope model with individual split-point times, values and curve form, and amplitude and frequency irregularities. The models parameters are perceptually relevant, and the model has previously been shown to have a high sound quality.
An important addition to the model, introduced in this paper, is the expression model, which enhances the use of the model in a musical context. The expression model models the behavior of the timbre attributes, when playing a different note, intensity, duration, or other style. It also introduces the vibrato and tremolo into the model. A major contribution is the notion of variants, in which the timbre is divided into an identity part, and an expression part. The variant is then a stochastic model of the deviation from the identity part of the sound, with a new instance at each new execution. Therefore, an interesting novelty is added to each new sound emitted, even if it's played with the same pitch, loudness and duration.

References

[1] Allen, J. B. *Short term spectral analysis, synthesis and modification by discrete fourier transform*. IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-25, No. 3, June. 1977.

[2] American Standard Association. Acoustical Terminology, New York, 1960.

[3] Andersen, T. H., K. Jensen. *Phase modeling of instrument sounds based on psycho acoustic experiments*, Workshop on current research directions in computer music, Barcelona, Spain, 2001.

[4] Auger F., P. Flandrin. *Improving the Readability of Time Frequency and Time Scale Representations by the Reassignment Method*, IEEE Transactions on Signal Processing, vol. 43, pp. 1068-1089, 1995.

[5] Backus, J. *The acoustical foundation of music*. John Murray Ltd. London, 1970.

[6] Beauchamp, J. *Synthesis by spectral amplitude and "Brightness" matching of analyzed musical instrument tones*. J. Acoust. Eng. Soc., Vol. 30, No. 6. 1982.

[7] Barrière, J-P (editor). *Le timbre, métaphore pour la composition*, C. Bourgois Editeur, IRCAM, 1991.

[8] Beauchamp, J. W. *Unix workstation software for analysis, graphics, modifications, and synthesis of musical sounds*, 94th AES Convention, preprint 3479, Berlin, Germany. 1993.

[9] Benade, A. H. *Fundamentals of musical acoustics.* Dover publications inc. New York. 1990.

[10] Bensa J., K. Jensen, R. Kronland-Martinet, S. Ystad. *Perceptual and Analytical Analysis of the effect of the Hammer Impact on the Piano Tones*, Proceedings of the ICMC, Berlin, Germany. 2000.

[11] Borum, S., K. Jensen. *Additive Analysis/Synthesis Using Analytically Derived Windows*, Proceedings of the DAFX, Trondheim, Norway, 1999.

[12] Bregman, A. S. *Auditory Scene Analysis*, The MIT Press, Massachussetts. 1990.

[13] Chaudhary, A. S. *Perceptual Scheduling in Real-time Music and Audio Applications*, Doctoral dissertation, Computer Science, University of California, Berkeley, CA. 2001.

[14] Chen, S. S., D. L. Donoho, M. A. Saunders. *Atomic decomposition by basis pursuit.* Dept. of Statistics Technical Report, Stanford University, February, 1996.

[15] Cheung, N-M., A. B. Horner. *Group synthesis with genetic algorithms*. J. Audio Eng. Soc. Vol. 44, No. 3, March. 1996.

[16] d'Alessandro, C., M. Castellengo. *The pitch of short-duration vibrato tones*, J. Acoust. Soc. Am, 95(3), pp. 1617-1630, March. 1994.

[17] Depalle Ph., T. Hélie. *Extraction of spectral peak parameters using a short time Fourier transform modeling and no sidelobe windows*. Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk, Oct. 1997.

[18] Desain, P., H. Honing. *Time functions function best as functions of multiple times.* Computer Music Journal, 16(2), 17-34, 1992.

[19] Ding, Y., X. Qian. *Processing of musical tones using a combined quadratic polynomial-phase sinusoid and residual (QUASAR) signal model*, J. Audio Eng. Soc. Vol. 45, No. 7/8, July/August 1997.

[20] Dubnov, S., N. Tishby, D. Cohen. *Investigation of frequency jitter effect on higher order moments of musical sounds with application to synthesis and classification*. Proc of the Int. Comp. Music Conf. 1996.

[21] Dubnov, S., X. Rodet. *Statistical modeling of sound aperiodicity.* Proc of the Int. Comp. Music Conf. 1997.

[22] Eaglestone, B., S. Oates. *Analytic tools for group additive synthesis.* Proc. of the ICMC, 1990.

[23] Fitz, K., L. Haken, P. Christensen. *Transient Preservation Under Transformation In an Additive Sound Model,* Proceedings of the International Computer Music Conference, Berlin, Germany. August 2000,

[24] Fitz, K., L. Haken. *Bandwidth enhanced modeling in Lemur*. Proc. of the ICMC, 1995.

[25] Fitz, K. Project: *Loris*. <http://sourceforge.net/projects/loris>, 08/10 2001.

[26] Fletcher, H. *Normal vibrating modes of a stiff piano string,* J. Acoust. Soc. Am., Vol. 36, No. 1, 1964.

[27] Fletcher, N. H., T. D. Rossing. *The physics of musical instruments*, Springer-Verlag. 1990.

[28] Freedman, M. D. *Analysis of musical instrument tones.* J. Acoust. Soc. Am. Vol. 41, No. 4, 1967.

[29] Friberg, A. *Generative rules for music performance: A formal description of a rule system.* Computer Music Journal, Vol. 15, No. 2, summer 1991.

[30] Fujinaga, I., K. MacMillan. *Realtime recognition of orchestral instruments*. Proc. of the ICMC. 2000.

[31] Gabrielson, A. *The performance of music*, *in* The psychology of music, D. Deutsch (editor), pp. 501-602, AP press, San Diego, USA, 2nd edition, 1999.

[32] Gordon, J. W. *The perceptual attack time of musical tones*. J. Acoust. Soc. Am. 82(2), July 1987.

[33] Grey, J. M. *Multidimensional perceptual scaling of musical timbres*. J. Acoust. Soc. Am., Vol. 61, No. 5, May 1977.

[34] Grey, J. M., J. A. Moorer. *Perceptual evaluation of synthesized musical instrument tones,* J. Acoust. Soc. Am., Vol. 62, No. 2, August 1977.

[35] Grey, J. M., J. W. Gordon. *Perceptual effects of spectral modification on musical timbres.* J. Acoust. Soc. Am. Vol. 63, No. 5, May 1978.

[36] Gribonval, R., P. Depalle, X. Rodet, E. Bacry, S. Mallat. *Sound signal decomposition using a high resolution matching pursuit*. Proc. ICMC, 1996.

[37] Guillemain, P. *Analyse et modélisation de signaux sonores par des représentations temps-frequence linéaires.* Ph.D. dissertation. Université d'Aix-Marseille II, 1994.

[38] Hajda, J. M., R. A. Kendall, E. C. Carterette, M. L. Harschberger. *Methodological issues in timbre research, in* The psychology of music, D. Deutsch (editor), pp. 253-306, AP press, San Diego, USA, 2nd edition, 1999.

[39] Harris, F. J. *On the use of windows for harmonic analysis with the discrete Fourier transform*. Proc IEEE, Vol. 66, No. 1, January 1978.

[40] Helmholtz, H. *On the Sensations of Tone,* reprinted in 1954 by Dover, New York, 1885.

[41] Herrera, P., J. Bonada. *Vibrato extraction and parameterization in the spectral modeling synthesis framework,* Proc. DAFx, Barcelona, Spain, 1998.

[42] Herrera, P., X. Amatriain, E. Batlle, X. Serra. *Towards Instrument Segmentation for Music Content Description: a Critical Review of Instrument Classification Techniques,* Proceedings of International Symposium on Music Information Retrieval, 2000.

[43] Hunt A., M. Wanderley. *Interactive Systems and Instrument Design in Music Working Group,* <http://www.notam.uio.no/icma/interactivesystems/wg.html>, October 4, 2001.

[44] Iverson, P., C. L. Krumhansl. *Isolating the dynamic attributes of musical timbre*. J. Acoust. Soc. Am. 94(5), November 1993.

[45] Jaffe, D. A., J. O. Smith. *Extension of the Karplus-Strong plucked-string algorithm*. Computer Music Journal, Vol. 7, No. 2, summer 1983.

[46] Järväinen, H., V. Välimäki, M. Karjalainen. *Audibility of inharmonicity in string instrument sounds, and implications to digital sound synthesis.* ICMC Proceedings, Beijing, China, 359-362. 1999.

[47] Jensen, K. *Envelope Model of Isolated Musical Sounds,* Proceedings of the DAFX, Trondheim, Norway, 1999.

[48] Jensen, K., G. Marentakis, *Hybrid Perception,* Papers from the 1st Seminar on Auditory Models, Lyngby, Denmark, 2001.

[49] Jensen, K., J. Arnspang. *Binary Tree Classification of Musical Instruments,* Proceedings of the ICMC, Beijing, China, 1999.

[50] Jensen, K. *Pitch Independent Prototyping of Musical Sounds,* Proceedings of the IEEE MMSP Denmark, 1999.

[51] Jensen, K. *The Control of Musical Instruments*, Proceedings of the NAM. Helsinki, Finland. 1996.

[52] Jensen, K. *Timbre Models of Musical Sounds*, PhD. Dissertation, DIKU Report 99/7, 1999.

[53] Kleczowski, P. *Group Additive Synthesis.* Computer Music Journal 13(1), 1989.

[54] Krimphoff, J., S. McAdams, S. Winsberg. *Caractérisation du timbre des sons complexes. II Analyses acoustiques et quantification psychophysique*. Journal de Physique IV, Colloque C5, Vol. 4. 1994.

[55] Kronland-Martinet, R. *The wavelet transform for analysis, synthesis, and processing of speech and music sounds*. Computer Music Journal, 12(4), 1988.

[56] Lindeberg, T. *Edge detection and ridge detection with automatic scale selection*, CVAP Report, KTH, Stockholm, 1996.

[57] Loscos, A., P. Cano, J. Bonada, M. de Boer, X. Serra. *Voice Morphing System for Impersonating in Karaoke Applications*, Proceedings of International Computer Music Conference 2000.

[58] Marchand, S. *Musical sound effects in the sas model*, Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99), NTNU, Trondheim, December 9-11, 1999

[59] Marchand, S. *Improving Spectral Analysis Precision with Enhanced Phase Vocoder using Signal Derivatives.* In Proc. DAFX98 Digital Audio Effects Workshop, pages 114--118, Barcelona, November 1998.

[60] Marentakis, G., K. Jensen. *Timbre Engine: Progress Report*, Workshop on current research directions in computer music, Barcelona, Spain, 2001.

[61] Marques, J. S., A. J. Abrantes. *Hybrid harmonic coding of speech at low bit-rates*, Speech Communication 14, 1994.

[62] Martínez J. M. *Overview of the MPEG-7 Standard (version 5.0)*, <http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm>, Singapore, March 2001.

[63] Matthews, M. V., J. E. Miller, E. E. David. *Pitch synchronous analysis of voiced speech*. J. Acoust. Soc. Am. Vol. 33, No. 2, February 1961.

[64] McAdams, S., J. W. Beauchamp, S. Meneguzzi. *Discrimination of musical instruments sounds resynthesized with simplified spectrotemporal parameters*, JASA 105(2), 1999.

[65] McAdams, S., S. Winsberg, S. Donnadieu, G. de Soete, J. Krimphoff. *Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes.* Psychological Research, 58, pp. 177-192. 1992.

[66] McAulay, R. J., T. F. Quatieri. *Speech analysis/synthesis based on a sinusoidal representation*, IEEE Trans. on Acoustics, Speech and Signal Proc., vol. ASSP-34, No. 4, August 1986.

[67] McIntyre, M. E., R. T. Schumacher, J. Woodhouse. *Aperiodicity in bowed-string motion*, Acustica, Vol. 49, 1981.

[68] Mellody, M., G. H. Wakefield. A model distribution study of the violin vibrato. Proc ICMC, 1997.

[69] Møller, A. *Akustisk guitar syntese.* Master Thesis, Computer Science Department, University of Copenhagen, 1996.

[70] Moorer, J. A. *The synthesis of complex audio spectra by means of discrete summation formulas*. J. Audio. Eng. Soc. Vol. 24, No. 9, November 1976.

[71] Moré, J. J. *The Levenberg-Marquardt algorithm: Implementation and theory.* Lecture notes in mathematics, Edited by G. A. Watson, Springer-Verlag, 1977.

[72] Pielemeier, W. J., G. H. Wakefield, M. H. Simoni. *Time-frequency analysis of musical signals*. Proc. of the IEEE, Vol. 84, No. 9. September 1996.

[73] Rayleigh, J. W. S. *The theory of sound*. reprinted in 1945 by Dover Publications. MacMillan company 1896.

[74] Recommendation ITU-R 85/10. *Methods for the subjective assessment of small impairments in audio systems, including multichannel sound systems.* International Telecommunication Union, Geneva, Switzerland. 16 March 1994.

[75] Richard, G., C. d'Allesandro. *Analysis, Synthesis and modification of the speech aperiodic component*, Speech Communication 19, 1996.

[76] Richard, G., C. d'Allesandro, S. Grau. *Musical noises synthesis using random formant waveforms.* Stockholm Music Acoustic Conference, Pub. of the royal Academy of Music, Stockholm, Sweden 1993.

[77] Risset, J-C. *Timbre analysis by synthesis: Representation, imitation and variants for musical composition*. *in* Representations of musical signals. G. de Poli, A. Piccialli, C. Roads, Editors. The MIT Press, London 1991.

[78] Röbel. A. *Adaptive Additive Synthesis of Sound*, Proceedings of the ICMC, Berlin, Germany, 1999.

[79] Rocchesso, D., F. Scalcon. *Bandwidth of perceived inharmonicity for physical modeling of dispersive strings*. *IEEE Transactions on Speech*

and Audio Processing, 7(5):597-601, September 1999.

[80] Rodet, X., P. Depalle, G. Poiret. *Speech analysis and synthesis methods based on spectral envelopes and voiced/unvoiced functions*. European Conference on Speech Technology, Edinburgh, September 1987.

[81] Rodet, X. *The additive analysis-synthesis package*, <http://www.ircam.fr/equipes/analyse-synthese/DOCUMENTATIONS/additive/index-e.html>, 08/10 2001.

[82] Rossignal, S. *Segmentation et indexation des signaux sonores musicaux*, Thèse de doctorat de l'université Paris VI, Paris, France 2001.

[83] Rovan, J. B., M. M. Wanderley, S. Dubnov, Ph. Depalle. *Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance*, Kansei- The Technology of Emotion Workshop - Genova - Italia, Oct. 3/4, 1997.

[84] Sandell, G. *Definitions of the word "Timbre"*, <http://sparky.parmly.luc.edu/sandell/sharc/timbredef.html>, 08/10 2001.

[85] Sandell, G. J., W. L. Martens. *Perceptual evaluation of principal-component-based synthesis of musical timbres*. J. Acoust. Soc. Am. Vol. 43, No. 12, December 1995.

[86] Schaeffer, P. *Traité des objets musicaux*, Editions de Seuil, 1966.

[87] Schumacher, R. T., C. Chafe. *Characterization of aperiodicity in nearly periodic signals*. J. Acoust. Soc. Am. Vol. 91 No. 1, January 1992.

[88] Schwarz, D., X. Rodet. *Spectral Envelope Estimation and Representation for Sound Analysis-Synthesis*, Proc. ICMC, pp 351-354, Beijing, China, 1999.

[89] Sekey, A., B. A. Hanson. *Improved 1-bark bandwidth auditory filter*. J. Acoust. Soc. Am. Vol. 75, No. 6, 1984.

[90] Serra, X.. J. Smith. *Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition*, Computer Music Journal, vol. 14, No. 4, winter 1990.

[91] Singh P. G., A. S. Bregman. *The Influence of different timbre attributes on the perceptual segregation of complex-tone sequences*, Journal of the Acoustical Society of America 1,02 1997.

[92] Slaney, M., M. Covell, B. Lassiter. 1996. *Automatic audio morphing*, Proc. IEEE Int. Conf. Acosut. Speech Signal Process. 2, 1001-1004, 1996.

[93] Strawn, J. *Orchestral Instruments: Analysis of performed transitions*, J. Audio. Eng. Soc, Vol. 34, No. 11, pp. 867-880. November 1986.

[94] Strong, W., M. Clark. *Synthesis of wind-instrument tones*. J. Acoust. Soc. Am. Vol. 41, No. 1, 1967.

[95] Tellman E., L. Haken, B. Holloway. *Timbre morphing of sounds with unequal numbers of features*. J. Audio Eng. Soc. Vol. 43, No. 9, September 1995.

[96] Truax, B. *Discovering inner complexity: Time shifting and transposition with a real-time granulation technique*. Computer Music Journal. 18(2), summer 1994.

[97] Verma, T. S., S. N. Levine, T. H. Y. Meng. *Transient Modeling Synthesis: A flexible analysis/synthesis tool for transient signals*, Proceedings of the ICMC, 1997.

[98] von Bismarck, G. *Timbre of steady sounds*. Acustica, Vol. 30, 1974.

[99] Vos, K., R. Vafin, R. Heusdens, W. B. Kleijn. *High quality consistent analysis-synthesis in sinusoidal coding*, AES 17th Conference High Quality Audio Coding, Florence, Italy, 1999.

[100] Wanderley M., M. Battier (eds). *Trends in Gestural Control of Music*. Ircam - Centre Pompidou, 2000.

[101] Wanderley, M. *Interaction musicien-instrument: Application au contrôle gestuel de la synthese sonore*. Thèse de doctorat de l'université Paris VI, Paris, France 2001.

[102] Warren, R. M. *Auditory Perception: A new analysis and synthesis*. Cambridge university press, Cambridge, 1999.

[103] Weinreich, G. *Coupled piano strings*, J. Acoust. Proc. Am., Vol 62, No. 6, pp. 1474-1484, December 1977.

[104] Wessel, D. *Timbre space as a musical control structure*, Computer Music Journal 3(29, pp.45-52, 1979.

[105] Zwicker, E., H. Fastl. *Psycho-acoustics: Facts and models*, 2nd edition, Springer-verlag, Heidelberg, Germany, 1999.