SOURCE-FILTER MODEL FOR QUASI-HARMONIC INSTRUMENTS

Henrik Hahn
Audio Communication Group,
Technische Universität Berlin
Berlin, Germany
henrik.hahn@campus.tu-berlin.de

Axel Röbel
Analysis/Synthesis Group,
IRCAM
Paris, France
axel.roebel@ircam.fr

*Juan José Burred**
Analysis/Synthesis Group,
IRCAM
Paris, France
burred@ircam.fr

Stefan Weinzierl
Audio Communication Group,
Technische Universität Berlin
Berlin, Germany
stefan.weinzierl@tu-berlin.de

ABSTRACT

In this paper we propose a new method for a generalized model representing the time-varying spectral characteristics of quasi - harmonic instruments. This approach comprises a linear source-filter model, a parameter estimation method and a model evaluation based on the prototype's variance. The source-filter-model is composed of an excitation source generating sinusoidal parameter trajectories and a modeling resonance filter, whereas basic-splines (B-Splines) are used to model continuous trajectories. To estimate the model parameters we apply a gradient decent method to a training database and the prototype's variance is being estimated on a test database. Such a model could later be used as a priori knowledge for polyphonic instrument recognition, polyphonic transcription and source separation algorithms as well as for resynthesis.

1. INTRODUCTION

The purpose of our approach is to define an accurate as well as compact representation of the time-varying spectral characteristics of a single, quasi-harmonic instrument sound. While we assume the spectral envelope to be determined by the partial's amplitude trajectories, our model is meant to predict the time-varying amplitude trajectories for signals of unknown origin. Therefore, in order to prototype an instrument, we need to estimate the model parameters using a training database and to evaluate the performance of each prototype, we use a test database to estimate the variance of the predicted partial's amplitude trajectories.

Two approaches for a compact representation of the spectral characteristics of quasi-harmonic instruments have been proposed recently. In [1] a representational model based on additive analysis and Principal Component Analysis (PCA) is presented in a first step, while in a subsequent stage, the spectral evolutions are modeled as Gaussian Processes, i.e., as trajectories of varying mean and covariance in PCA space. Applied to musical instrument recognition, the model has been shown to significantly improve classification results compared to a Mel-Frequency-Cepstral-Coefficient based method. A source-filter-decay model is proposed in [2] and successfully applied to musical content analysis in [3] and [4]. In that approach the spectral envelope of an instrument sound is modelled by a source representing a vibrating object and a resonance filter related to the instrument's body which colors the generated sound. A decay filter is further used to model the time-varying characteristics.

* J.J. Burred is now with Audionamix, Paris, France.

In our approach we also adopt a linear source-filter model with similar interpretations of its components, but we extend this approach by taking the time variability into account for the complete amplitude envelope including attack and release regions. We will further introduce BSplines [5] for modeling the time varying spectral envelope as a smooth trajectory with respect to the different regions of the amplitude envelope as well as for modeling the filter's resonance curve. Finally, this yields a model parameterization determined only by the definition of the B-Splines.

In section 2 we will give a comprehensive description of our assumed signal model as well as our proposed source-filter model, while section 3 describes how to estimate the model parameters using a training database. Section 4 will represent the variance measure and in section 5, results for some selected prototypes including their variance values and resynthesized amplitude trajectories are presented.

2. THE MODEL

Based on the general assumption that the spectral characteristics of an instrument sound are being determined by the partial's amplitude trajectories, we start with an overview of the signal model being used throughout this work. The subsequent paragraphs will show how this signal model will be represented by our approach for a source-filter-model and how resynthesis can be employed.

2.1. Signal Model

In additive analysis/synthesis it is assumed that a signal $x[n]$ can be approximated as a sum of quasi-stationary sinusoids [6], so called partials.

$$x[n] \approx \tilde{x}[n] = \sum_{k=1}^{K} a[k,n] cos(\phi[k,n]) \tag{1}$$

In equation (1) k is the partial index, while K denotes the amount of partials, a denotes the amplitude for partial k at time frame n, as ϕ its phase. Though, the signal is modeled by its deterministic component only. We furthermore scale the partial amplitude values by their summed energy maximum over time to analyse the signal characteristics independently of the actual energy, yielding a normalized maximum signal level of 0dB. As a consequence, we denote $A[k, n]$ to be the scaled energy level of a partial's amplitude given in dB.

Since the spectral envelope varies over time, particularly during attack and release, an assumption has to be made, with regard to this variability. As time itself is an unfavorable unit due to varying durations of attack and release and by arbitrary signal lengths, we assume the variation of the spectral envelope to be directly related to the relative energy level of the signal. Accordingly, we assume the spectral envelope to be constant for a specific relative energy level and consider different envelopes for different levels. This also includes the assumption of the spectral envelope to be independent of the actual volume. Thus, the relative energy level $L[n]$ over time is given by 3

$$E[n] = \sum_{k=1}^{K} (a[k, n])^2 \tag{2}$$

$$L[n] = 10 \cdot log_{10} \left(\frac{E[n]}{\max(E[n])} \right) \tag{3}$$

Moreover, we have to take into consideration that levels below 0dB may correspond to either the attack or release, but spectral envelopes may differ for these regions. We therefore have to determine the signal's attack n_A and release n_R time frames and find some suitable partitioning n_a, respective n_r of an entire signal $\tilde{x}[n]$. In case of an continuously excited signal, we assume a sustain part to be present in the signal and therefore n_a denotes time frames, covering the attack and parts of the sustain within the signal, whereas n_r denotes parts of the sustain to full release region. For an impulsive excited signal in contrast, n_a denotes the attack region only as n_r does for the release region, because a sustain part is assumed to be absent. To determine n_A and n_R we use a simple threshold method applied to the relative energy level over time and distinguish between the continuous and impulsive case by applying different values for the threshold denoted by γ. The continuous case is shown in 1 and a suitable partitioning using adjoint bounds to determine n_a and n_r is presented in the inequalities 4 and 5. Here we use a threshold value below 0dB, whereas in the impulsive case the threshold is set to 0dB giving $n_A = n_R$ reflecting the absence of a sustain region.

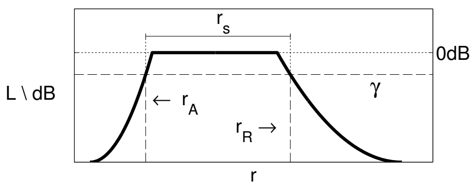

Figure 1: *Symbolic representation for attack/release determination in the case of a continuous excitation.*

$$r_a : r \leq \frac{1}{2}(r_A + r_R) \tag{4}$$

$$r_r : r \geq \frac{1}{2}(r_A + r_R) \tag{5}$$

Regarding our signal model the partitioned amplitudes of the partials can be denoted $A[k, n_a]$ and $A[k, n_r]$. The resulting amount of time frames for either the attack to sustain or sustain to release regions will later be referred to by N_A and N_R.

Additionally, as we are only considering quasi-harmonic instruments, the frequency values of the partials will be approximated as being in an integer ration regarding its fundamental and being constant throughout an entire signal leading to eq. 6

$$f(k) = f_0 \cdot k \quad , \quad k = 1 \ldots K \tag{6}$$

While f_0 denotes the fundamental frequency, $f(k)$ gives a sequence of frequency values of size K. As a result this approximation significantly simplifies our modeling approach.

2.2. Source-Filter Model

Our approach is based on the distinction of features being correlated to the fundamental frequency f_0 and features being independently of the fundamental. Features correlated to f_0 may refer to characteristics such as odd harmonics being stronger than even ones and therefore are better described as a function of the partial index k instead of actual frequencies. In contrast, formants or resonances refer to f_0 independent features and have to be described explicitly by their frequency values. In our source-filter model we refer to this distinction by expressing the f_0 correlated features within the source and the f_0 independent features within the filter. By this approach, the source will generate an envelope as a function of the partial index and without considering the fundamental, while the filter colors this envelope by taking the frequencies of the partials into account explicitly.

2.2.1. Source Model

By assuming the source to include the f_0 correlated features we use an oscillator model to reflect this. Additionally, in contrast to [2], we assume the variation of the spectral envelope in time to be correlated with the fundamental frequency, rather than independent from f_0. Thus, the temporal behaviour of the spectral envelope is assumed to be related to the partial index rather than to actual frequencies. This makes our oscillator dependent on the relative signal level L as well as on the partial index k. By taking into account that the progression of each partial over the relative energy is continuous, we model the partial's trajectories using piecewise polynomials. As described in [5], the linear superposition of weighted basic-splines (B-Splines) gives maximally smooth trajectories and the B-Spline functions are completely determined by the size of their segments and their order o, denoting the amount of segments covered by a single B-Spline polynomial. Due to linear superposition, the order of the piecewise polynomials follows $o - 1$, therefore the order o also defines the degree of smoothness of the function approximated by the superposition. As B-Spline polynomials are defined to converge to zero at their limits, zero size segments are used to model trajectory values at the limits differing from 0. Figure 2 shows a set of 7 B-Splines U_p as functions of level L. So, as we want to model the spectral envelope for a specific range of relative levels of an entire signal, we need

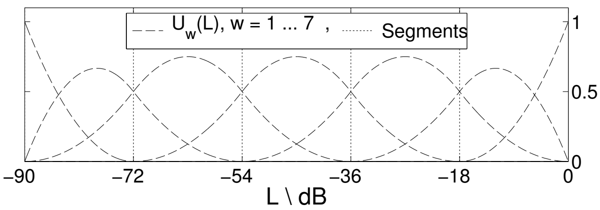

Figure 2: *B-Spline polynomials U_p of order 3 for 5 segments over a level range of -90 to 0dB. Two zero size segments have been added to both extrema.*

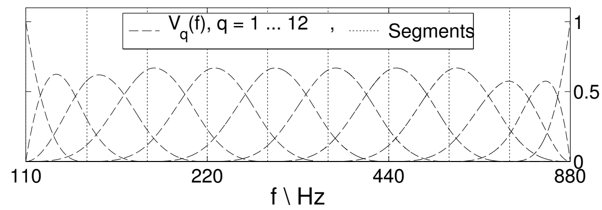

Figure 3: *B-Spline polynomials V_q of order 4 for 3 octaves with a segment size of $1/3$ octave. 3 zero size segments are added at extreme values.*

to introduce two different oscillators to either express the signals's attack to sustain or sustain to release regions.

$$O(k,L)_A = \sum_p^P g_{k,p}^A U_p(L) \tag{7}$$

$$O(k,L)_R = \sum_p^P g_{k,p}^R U_p(L) \tag{8}$$

In equations 7 and 8 the source model for the attack to sustain and the sustain to release oscillator is shown – indicated by the subscripts A and R, respectively. Each oscillator value is expressed by the weighted superposition of B-Spline polynomials for a single partial k at an arbitrary level L. Therefore, for both oscillators the weighting paramters g represent a sequence of coefficients for the piecewise polynomials for each single partial. The overall number of polynomials is denoted by P. Hence, the source generates an entire partial envelope for arbitrary signal levels, whereas the progression of a single over the relative level is expressed as a continuous trajectory. This perspective holds for the attack to sustain as well as the sustain to release oscillator.

2.2.2. Filter Model

The filter covers every part of an instrument not directly associated with the source. This primarily relates to the instrument's body. As we want the filter to model the features independent of f_0, the filter is assumed to lower or raise the partial's amplitude values $O(k,L)$ excited by the source with respect to their actual frequency values $f(k)$. Furthermore, the filter is assumed to be time-invariant and therefore frequency dependent only. Since we are only using information regarding the partials for all instrument signals, all information regarding the filter's frequency response will only be obtained at their frequency positions. But, as with the partial trajectories generated by the oscillators, the frequency response is assumed to be continuous, thus we also use B-Splines to model the filter. In contrast to the oscillator models, the B-Spline functions have to be defined in the frequency domain and as we expect the filter curve to exhibit prominent resonance peaks at lower levels and less prominent but dense peaks at higher frequencies, we propose frequency dependent segment sizes. Therefore, the segments will be distinguished by multiples of octave bandwidths with factors less or equal to 1, starting with the lowest possible fundamental frequency of each instrument. Consequently, the filter may model resonances and formants at lower frequencies more accurately and averaged within certain bandwidths at higher frequencies. Figure 3 shows 12 B-Spline polynomials V_q as a function of frequency f. So we can finally express the filter as a weighted, linear superposition of B-Splines for a complete set of polynomials denoted by Q.

$$F(f) = \sum_q^Q z_q V_q(f) \tag{9}$$

The weights z_q therefore denote a sequence of size Q, weighting the appropriate B-Spline functions. Note, that the actual amount of B-Splines will later be determined by the bandwidth expressed as a fraction of an octave that will fit into the frequency range beginning the the lowest possible fundamental of each instrument up to some maximum frequency value.

2.2.3. Synthesis

As our model approach is meant to predict the time-varying spectral envelope for a signal of unknown origin, the synthesis has tobe done with respect to this signal, considering its relative level L and fundamental frequency f_0. Therefore, these informations have to be determined in advance. Because $O(k,L)_A$ and $O(k,L)_R$ depend on the level L, the relative level of the signal has to be resolved as a function of n_a and n_r, respectively, utilizing eq. 3 and as $F(f)$ depends on the frequency, the fundamental frequency f_0 has to be determined to solve eq. 6. Taking this into account, prediction of the time-varying spectral envelope is given by a linear combination of the appropriate oscillator model and the filter shown in eq. 10 with respect to the level function $L[n_a]$ and $f(k)$.

$$\hat{A}(k, L[n_a], f_0)_A = O(k, L[n_a])_A + F(f_0 \cdot k) \tag{10}$$

Note, eq. 10 only shows the synthesis for the attack to sustain regions as conversion for the sustain to release region can easily be done by using the oscillator $O(k,L)_R$ and level function $L[n_r]$.

As n_a and n_r are adjoint within the temporal development of the signal, the spectral envelope predictions employed by the synthesis formula for both oscillator models can easily be connected to create the final time-varying spectral envelope.

3. PARAMETER ESTIMATION

Parameter estimation has to be done jointly to the oscillator models as well as to the filter model and is applied to the weighting coefficients of the B-Spline functions only. This introduces a model parameterization that is restricted to the amount of segments over a predefined level range for the oscillator models as well as the size of the segments for the filter model given as a fraction of octave bandwidths and their respective orders. Since we model the partial's amplitude trajectories according the relative signal level of the signal, it is obvious to determine $U_p(L[n_a])$ and $U_p(L[n_r])$ as well as $V_q(f(k))$ in advance, because while parameter estimation, these three functions will remain constant for each single training sample and can further be regarded as a projection of the training samples from input space to model space. For convenience and better readability, we introduce a matrix/vector notation, shown in table 1. The parameters to estimate are shown at the top, while

$\mathbf{G}_A$	$[K \times P]$	:	$g^A_{k,p}$
$\mathbf{G}_R$	$[K \times P]$	:	$g^R_{k,p}$
$\mathbf{z}_A$	$[Q \times N_A]$	:	z_q
$\mathbf{z}_R$	$[Q \times N_R]$	:	z_q
$\mathbf{U}_A$	$[P \times N_A]$	:	$U_p(L[n_a])$
$\mathbf{U}_R$	$[P \times N_R]$	:	$U_p(L[n_r])$
$\mathbf{A}_A$	$[K \times N_A]$	:	$A(k, n_a)$
$\mathbf{A}_R$	$[K \times N_R]$	:	$A(k, n_r)$
$\mathbf{V}$	$[Q \times K]$	:	$V_q(f(k))$
$\hat{\mathbf{A}}_A$	$[K \times N_A]$	:	$\hat{A}(k, n_a, f_0)_A$
$\hat{\mathbf{A}}_R$	$[K \times N_R]$	:	$\hat{A}(k, n_r, f_0)_R$

Table 1: *Matrix/Vector notation conventions*

all data dependent variables are shown at the middle and the two predicted spectral envelopes are shown at the bottom. Note, as we assume the partial's frequencies to be constant over time frames n, the matrices $\mathbf{z}_A$ and $\mathbf{z}_R$ will also be invariant over N_A and N_R.

Since we make use of a gradient decent method to estimate the model parameters, we need to define a cost function and its gradients. In this method, the parameters $\mathbf{G}_A$, $\mathbf{G}_R$ and $\mathbf{z}$ will adapt iteratively according to their negative gradient of the cost function until the gradient function converges. Finally, after the cost function has converged, the fixed set of model parameters is called a prototype for the instrument, which has been used for estimation.

3.1. Cost Function

For estimation of the model parameters we introduce a squared cost function. As shown in eq. 11, the Forbenius norm is used to indicate that we are taking the entrywise squared values of all values within the matrix and average over all partials k and time frames n to resolve the equation to a scalar value. Finally, both costs for the attack to sustain and sustain to release are averaged.

$$c = \frac{1}{2K}\Big(\frac{1}{2N_A}\Big|\Big|\Big(\mathbf{G}_A\mathbf{U}_A + \mathbf{V}^T\mathbf{z}_A\Big) - \mathbf{A}_A\Big|\Big|_2^2 + \frac{1}{2N_R}\Big|\Big|\Big(\mathbf{G}_R\mathbf{U}_R + \mathbf{V}^T\mathbf{z}_R\Big) - \mathbf{A}_R\Big|\Big|_2^2\Big) \tag{11}$$

Since this equation gives the cost for a single training sample, we average over all sample cost to measure the cost for the complete training database.

3.2. Gradient Functions

To get the gradients, the first derivative of the cost function with respect to the parameters has to be solved. This can be done by applying the chain rule once.

$$\frac{\partial c}{\partial \mathbf{G}_A} = \frac{1}{N_A}\big((\mathbf{G}_A\mathbf{U}_A + \mathbf{V}^T\mathbf{z}) - \mathbf{A}_A\big)\mathbf{U}_A^T \tag{12}$$

$$\frac{\partial c}{\partial \mathbf{G}_R} = \frac{1}{N_R}\big((\mathbf{G}_R\mathbf{U}_R + \mathbf{V}^T\mathbf{z}) - \mathbf{A}_R\big)\mathbf{U}_R^T \tag{13}$$

Note, as we want to get the gradients for all K sequences of B-Spline coefficients for the oscillator models, neither averaging over k nor averaging over the oscillator models has to be done. For the filter coefficients, on the other hand, averaging over both oscillators remains necessary as well as averaging over all time frames, deploying its time invariance. Therefore, only a single gradient vector $\mathbf{z}$ of size Q has to be resolved.

$$\frac{\partial c}{\partial \mathbf{z}} = \frac{1}{2K}\Bigg(\mathbf{V}\frac{1}{N_A}\sum_{n_a}^{N_A}\Big(\big(\mathbf{G}_A\mathbf{U}_A + \mathbf{V}^T\mathbf{z}_A\big) - \mathbf{A}_A\Big) + \mathbf{V}\frac{1}{N_R}\sum_{n_r}^{N_R}\Big(\big(\mathbf{G}_R\mathbf{U}_R + \mathbf{V}^T\mathbf{z}_R\big) - \mathbf{A}_R\Big)\Bigg) \tag{14}$$

4. MODEL VARIANCE

To estimate the model's prediction accuracy, we employ a statistical model shown in eq. 15, proposing the true spectral envelope A being determined by our predicted envelope $\hat{A}$ and some additive noise R. The noise is assumed to be drawn from a gaussian distribution and independent of L and f_0.

$$A[k, n] = \hat{A}(k, L[n], f_0) + R[k, n] \tag{15}$$

Therefore, by turning our regression model into a statistical, its variance is determined by the variance of the additive noise. Additionally, this variance may also be interpreted as the mean square error or our estimator $\hat{A}$. In contrast to all previous considerations, we will estimate the model's variance using linear amplitude values, indicated by their respective lower case letters. Eq. 16 shows how to estimate the variance, whereas the level contour $L[n]$ and fundamental frequency f_0 of the data sample $a[k, n]$ has been used to predict the spectral envelope $\hat{a}(k, L[n], f_0)$.

$$\sigma^2 = \frac{1}{KN}\big|\big|a[k, n] - \hat{a}[k, L[n], f_0]\big|\big| \tag{16}$$

We take the expectation of all variance values of a complete test database, which gives us a single scalar value to quantify the model's prediction accuracy and finally transform it to decibel scale.

5. RESULTS

We evaluated our approach using 7 instruments taken from the RWC musical instrument database [7]. This database contains three variants for each single instrument, each played by a different instrumentalist and various dynamic styles, giving us the possibility to achieve a high degree of generalization for the expected spectral envelope. Various model parameterizations regarding the definition of the B-Splines for the oscillator models as well as the filter model have been applied and for each a 10-fold cross-validation method has been used. Since online estimation has been shown to perform up to 5 times faster than its offline counterpart, by means of the number of iterations needed for convergence of the cost function, it has been chosen to be the favorable estimation method. Furthermore, the number of iterations being needed for convergence can significantly be reduced by incorporating a priori knowledge about the characteristics of the filter as well as the oscillators while initialization of the weighting coefficients. Nethertheless, even random initialization gives comparable results apart from the number of iterations needed. The results presented here are selected in terms of the minimum cross-validation error for all applied parameterizations.

For visualization of the prototypes we use the level sequence $L^{sigma} = \{-30, -27.5, \ldots, -2.5, 0\}$ to generate the respective partial envelopes within our source models, whereas for the sustain to release oscillator model, the sequence is resolved in reversed order. Each single level value therefore reflects a partial envelope, indicated as a single dotted line within the oscillator models. In figures 4, 5 and 6 prototypes for the clarinet, the grand piano as well as the violin are shown. At the top of each figure, the partial envelopes generated by the oscillators are shown for all values of L^{σ} with respect to the partial index k. Only the first 16 partials are presented as they carry most of the signals charcteristics. The filter is band limited to low frequencies by the lowest possible fundamental of each instrument and to high frequencies by half of the sampling rate. As can be seen in all three figures, various f_0 independent resonances and formants have been estimated by the filter as well as f_0 correlated features have been estimated by the oscillator models. Moreover, the time-varying spectral envelope has also been reflected by the varying partial envelopes of the oscillator models. The variance measures used to quantify the model's prediction performance is shown in table 2. As can be seen, the variance values are quite close to each other, beside the value for the grand piano's prototype which has an much lower value. All values indicate a low average variance for the additive noise introduced by our model and therefore spectral envelope predictions close actual measured ones. This can also bee seen in the synthesis examples. Fig. 7 shows the measured partial amplitude trajectories $A[k, n]$ for a clarinet playing an Bb4 on the

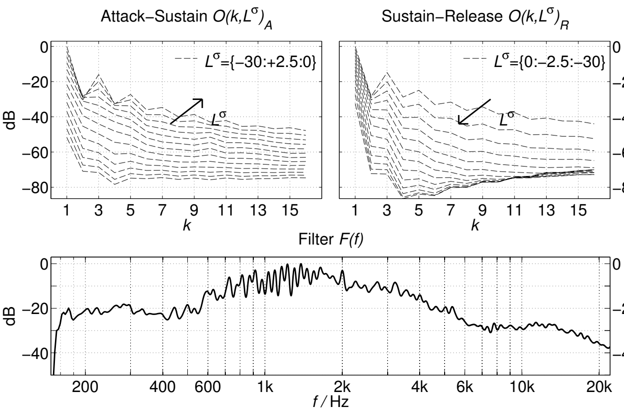

Figure 4: Prototype of a clarinet using 5 B-Spline segments for the oscillators and a segment size of $\frac{1}{24}$ octave for the B-Spline modeling of the filter. All B-Spline orders are set to 3.

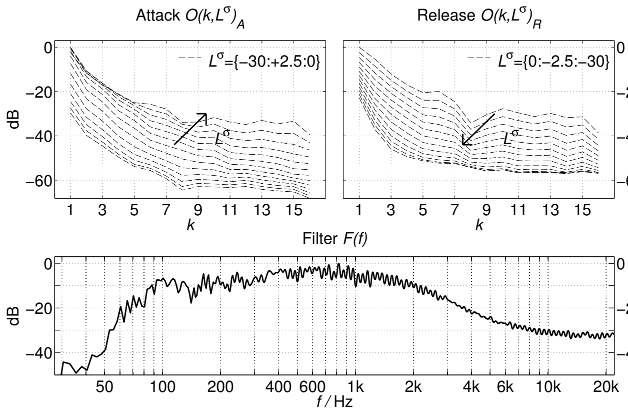

Figure 5: Prototype of a grand piano using 5 B-Spline segments for the oscillators and a segment size of $\frac{1}{24}$ octave for the B-Spline modeling of the filter. All B-Spline orders are set to 3.

Instrument		σ^2_{dB}
Trumpet	:	-25.5
Alto Saxophone	:	-26.5
Clarinet	:	-26
Oboe	:	-24
Piano	:	-34.5
Violin	:	-26
Violin Cello	:	-29

Table 2: Variance values in dB for the selected 7 instrument prototypes

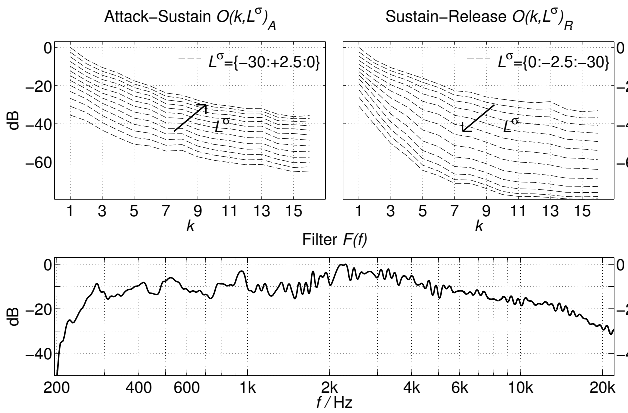

Figure 6: Prototype of a violin using 5 B-Spline segments for the oscillators and a segment size of $\frac{1}{24}$ octave for the B-Spline modeling of the filter. All B-Spline orders are set to 3.

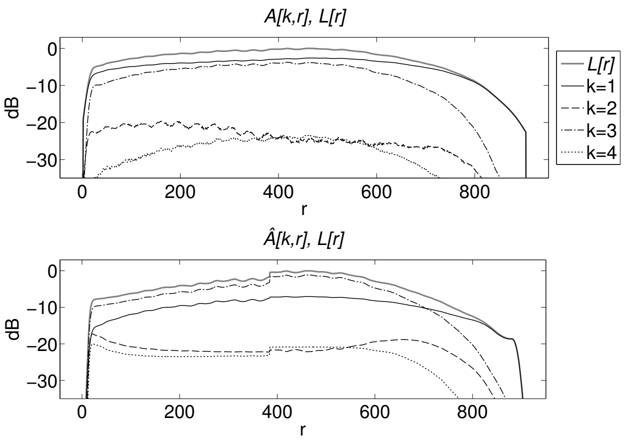

Figure 7: *Original (top) and predicted (bottom) amplitude trajectories for the first 4 partials of a clarinet playing Bb4*

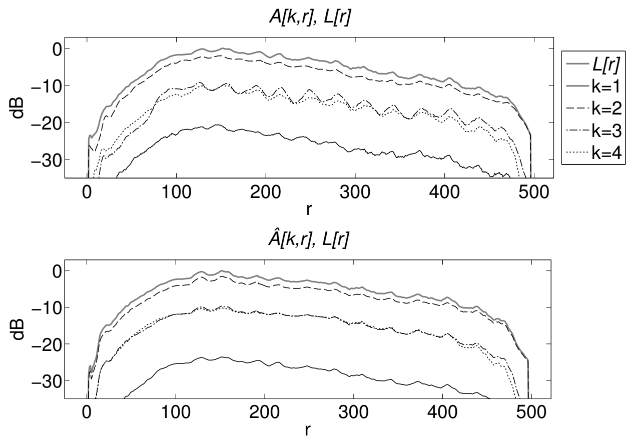

Figure 9: *Original (top) and predicted (bottom) amplitude trajectories for the first 4 partials of a violin playing Ab3*

top and the corresponding predicted partial's amplitude trajectory $\hat{A}(k, L[n], f_0)$ of our prototype of a clarinet on the bottom. Fig.

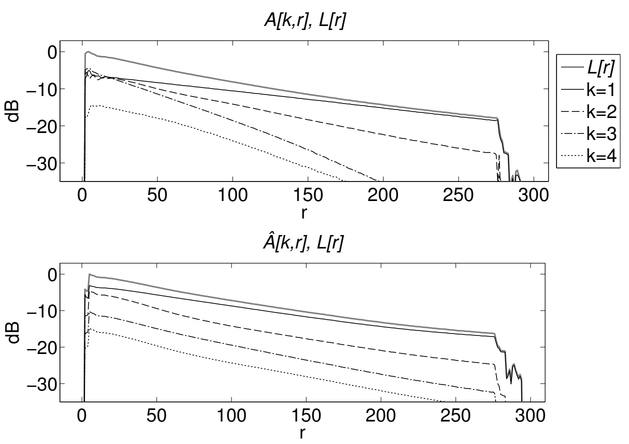

Figure 8: *Original (top) and predicted (bottom) amplitude trajectories for the first 4 partials of a grand piano playing F3*

8 presents the respective trajectories for a violin playing Db5 and fig. 9 shows the partial's amplitudes of a piano playing an E1. As can bee seen in the three figures, the prototypes predict the partial's amplitude trajectories with high accuracy.

6. CONCLUSION

In this paper we have shown a new approach for representing quasi-harmonic instruments by a linear source-filter model with the possibility to predict the time-varying spectral characteristics of an instrument with respect to the fundamental frequency and the relative level of the signal. We have further given the mathematical basis to estimate the model parameters using a trainings database and to estimate the variance of such a prototype. As shown by our results, the selected prototypes have estimated f_0 correlated as well as f_0 independent features. Even though, we believe our results are promising, we will improve the learning algorithm by utilizing a conjugate gradient method to increase the models performance regarding its statistical variance and in future research we will apply our model to music information retrieval and source separation algorithms.

7. REFERENCES

[1] Juan José Burred, Axel Röbel, and Xavier Rodet, "An accurate timbre model for musical instruments and its application to classification," in *First Workshop on Learning the Semantics of Audio Signals*, Athens, Greek, December 2006.

[2] Anssi Klapuri, "Analysis of musical instrument sounds by source-filter-decay model," *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 1, pp. I–53–I–56, 2007.

[3] Tuomas Virtanen and Anssi Klapuri, "Analysis of polyphonic audio using source-filter model and non-negative matrix factorization," in *TODO (AMAC)*, 2006.

[4] Toni Heittola, Anssi Klapuri, and Tuomas Virtanen, "Musical instrument recognition in polyphonic audio using source-filter model for sound separation," in *10th International Society for Music Information Retrieval Conference (ISMIR)*, 2009, pp. 327–332.

[5] Axel Röbel, "Adaptive additive modeling with continuous parameter trajectories," *IEEE Transactions on Speech and Audio Processing*, vol. 14-4, pp. 1440–1453, 2006.

[6] X. Serra, *Musical Signal Processing*, chapter Musical Sound Modeling with Sinusoids plus Noise, pp. 91–122, G. D. Poli, A. Picialli, S. T. Pope, and C. Roads Eds. Swets & Zeitlinger, Lisse, Switzerland, 1996.

[7] Masataka Goto and Takuichi Nishimura, "Rwc music database: Music genre database and musical instrument sound database," in *4th International Society for Music Information Retrieval Conference (ISMIR)*, 2003, pp. 229–230.

[8] Julius O. Smith III, *Spectral Audio Signal Processing*, online publication, accessed March 2010, Available at http://www.dsprelated.com/dspbooks/sasp/.