HIGH-RESOLUTION PARAMETRIC MODELING OF STRING INSTRUMENT SOUNDS

Matti Karjalainen and Tuomas Paatero

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P.O. Box 3000, FI-02015, TKK, Finland
phone: +358 9 451 2490, email: matti.karjalainen@.hut.fi
web: www.acoustics.hut.fi

ABSTRACT

Parametric modeling of musical instrument sounds is useful because it makes possible to represent signals in a compact form, to resynthesize them in a modified or morphed way, or to apply the parameters in physical and perceptual studies of acoustic instruments. In this paper we propose high-resolution pole-zero modeling techniques that are applicable to parametric representation of string instrument sounds, assuming approximately linear and time-invariant behavior of the instrument. An example of applying the techniques to the acoustic guitar sound is presented.

1. INTRODUCTION

Parametric models of musical instrument sounds are highly useful in audio, computer music, and music acoustics research. A parametric model allows for compact representation of signals in analysis and synthesis (including modification and morphing) as well as for the understanding of related phenomena in basic research.

There are different techniques available for parametrization of musical instrument signals. For example sinusoidal (or spectral) modeling is a general methodology, applicable to any sound composed of discrete spectral components. However, its parameters may not be explicitly related to underlying physical phenomena. Sinusoidal modeling also requires additional methods to cope with transients that are typical in the attacks of instrument sounds.

When instruments under study contain resonators that are approximately linear and time-invariant, a source-filter model can represent its operation by an excitation (source signal) and a system function (filter). Any such system can be specified by poles and zeros in the transfer function, whereby complex conjugate pole pairs are typically related to modes and resonances of the system.

Plucked and struck string instruments fit well to this scheme. For example in the guitar the string is excited by a brief impulse-like plucking event, after which the autonomous vibration modes decay exponentially. The source-filter approach using excitation and all-pole or pole-zero filtering is a natural solution in such cases. For synthesis purposes, special filter structures such as digital waveguides (DWG) [1] can be computationally more efficient than generic pole-zero filters. Even then, a general filter model can be used as an intermediate step in the calibration of the DWG models.

Generic pole-zero modeling of string instrument sounds is not simple, however. Difficulties come from the high order of the filter models required and the positions of complex poles very close to the unit circle in the z-domain. High-resolution techniques are needed to solve the filter parameters.

The problem domain of this paper is to elaborate high-resolution pole-zero methods for the modeling of string instrument sounds. We first concentrate on finding accurate values for the poles that represent modal behavior. An analysis method called FZ-ARMA is used for finding these poles. Then the final pole-zero filter and the excitation is achieved by Kautz filter models. As a case study, the approach is applied to the modeling of the acoustic guitar sound.

2. HIGH-RESOLUTION MODE ANALYSIS

If the response of a string to excitation were an AR (autoregressive) process, it could be analyzed by straightforward AR modeling techniques, such as the linear prediction (LP) [2], resulting in an all-pole filter. This is not manageable in practice because the pole positions close to the unit circle makes the analysis numerically too critical, and the allocation of the number of poles for each partial of the string cannot be easily controlled. It is important to notice that the two polarizations of string vibration with slightly different frequencies make each partial to be a sum of two decaying sinusoids, exhibiting beating or two-stage decay of the envelope [3], thus needing at least two pole pairs to represent it properly.

Since the string response to excitation is not a minimum-phase signal, the AR modeling approach is even in theory a wrong tool. ARMA (autoregressive moving average) models are capable to match such responses, but due to iterative solutions, models that are higher in order than 20–200 may not converge to a stable and useful result.

2.1 FZ-ARMA analysis

To avoid the problems in resolution and computational precision discussed above, we have developed a subband technique called FZ-ARMA (frequency-zooming ARMA) analysis [4, 5]. Instead a single model, global over the entire frequency range, the signal is pole-zero modeled in subbands, i.e., by zooming to a small enough band at a time, thus allowing a filter order low enough and individually selectable to each subband. This helps in resolving resonant modes that are very sharp or close to each other in frequency. The FZ-ARMA analysis consists of the following steps.

(i) Select a frequency range of interest, e.g. a few Hz wide frequency region around the spectral peak of a partial.

(ii) Modulate the target signal (shift in frequency by multiplying with a complex exponential) to place the center of the frequency band, defined in (i), at the origin of the frequency axis by mapping
$$h_m(n) = e^{-j\Omega_m n} h(n) \qquad (1)$$
where $h(n)$ is the original sampled signal, $h_m(n)$ the down-modulated one, n is the sample index, and Ω the (normalized) modulation frequency. This rotates the poles of transfer function by $\Omega_{i,rot} = \Omega_i - \Omega_m$.

(iii) Apply lowpass filtering to the complex-valued modulated signal in order to attenuate its spectral content outside the zoomed band of interest.

(iv) Decimate (down-sample) the lowpass filtered signal according to its new bandwidth. This zooms system poles z_i by
$$z_{i,zoom} = z_i^{K_{zoom}} = |z_i|^{K_{zoom}} e^{j(\Omega_i - \Omega_m) K_{zoom}} \qquad (2)$$
where K_{zoom} is the zooming factor, and $z_{i,zoom}$ are the mapped poles in the zoomed frequency domain.

(v) Estimate an ARMA (pole-zero) model for the obtained decimated signal in the zoomed frequency domain. For this we have applied the iterative Steiglitz-McBride algorithm (function `stmcb.m` in Matlab).

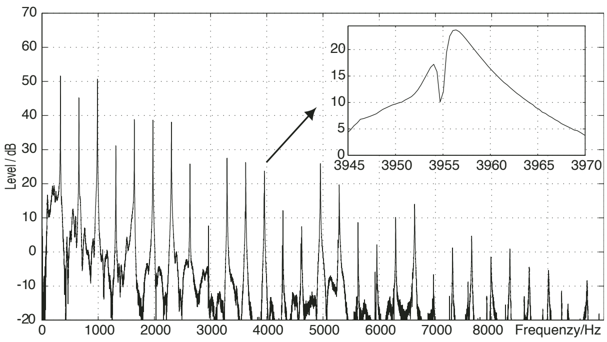

Figure 1: Spectrum of plucked string sound of a classical acoustic guitar, open string E_4. Spectrum zoomed to a single partial is shown in the subplot.

(vi) Map the obtained poles back to the original frequency domain by operations inverse to the above-presented ones. Zeros cannot be utilized as easily, thus we don't use them in this study for the final modeling. There may also be poles that correspond to the truncated frequency band edges, thus needing to be excluded. Therefore only relevant poles are directly useful parameters.

When applying pole-zero modeling, the selection of the number of poles has to be made appropriately according to the characteristics of the problem. The number of poles[1] should correspond to the order of the resonator to be modeled. For example a partial ('harmonic') of string vibration is composed of two polarizations, thus the partial may exhibit more than one peak in the frequency domain and beating or two-stage decay in the time-domain envelope. Figure 1 depicts the spectrum of a plucked guitar sound with a subplot zoomed into one partial having two spectral peaks.

A proper number of zeros in FZ-ARMA modeling is needed to make it able to fit the phases of the decaying sinusoids as well as modeling of the initial transient. Often this number is not very critical, and it can be somewhat higher than the number of poles.

The zooming factor K_{zoom} can be selected so that the analysis bandwidth contains most of the energy of the resonances to be modeled, keeping the order (number of poles) manageable.

3. POLE-ZERO MODELING BY KAUTZ FILTERS

The *Kautz filter* inherits its name due to a rediscovery in the early signal processing literature [7, 6] of an even older mathematical concept related to rational representations and approximations of functions [8]. The generic form of a Kautz filter is given by the transfer function

$$\begin{aligned} \hat{H}(z) &= \sum_{i=0}^{N} w_i G_i(z) \\ &= \sum_{i=0}^{N} w_i \left(\frac{\sqrt{1 - z_i z_i^*}}{1 - z_i z^{-1}} \prod_{j=0}^{i-1} \frac{z^{-1} - z_j^*}{1 - z_j z^{-1}} \right), \end{aligned} \quad (3)$$

where w_i, $i = 0, \ldots, N$, are somehow assigned tap-output weights. The orthonormal Kautz functions $G_i(z)$, $i = 0, \ldots, N$, are determined by any chosen set of stable poles: $\{z_j\}_{j=0}^{N}$, such that $|z_j| < 1$. Figure 2 is hopefully more instructive than formula (3). The time-domain counterpart of (3) is $\hat{h}(n) = \sum_{i=0}^{N} w_i g_i(n)$, where functions $\{g_i(n)\}$ are impulse responses or inverse z-transforms of functions $\{G_i(z)\}$. The meaning of orthonormality is specified, e.g., by using the time-domain inner products, $(g_i, g_k) := \sum_{n=0}^{\infty} g_i(n) g_k^*(n) = 0$ for $i \neq k$, and $(g_i, g_i) = 1$.

[1]Only the poles with positive imaginary component are used in FZ-ARMA analysis, representing the complex-conjugate pole pairs needed for real-valued filter model.

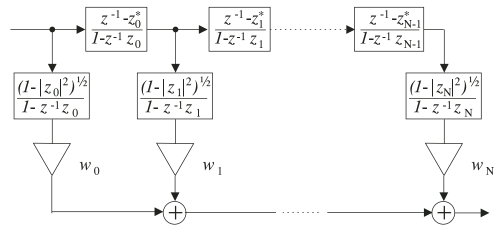

Figure 2: The Kautz filter. For $z_i = 0$ in (3) it degenerates to an FIR filter, for $z_i = a$, $-1 < a < 1$, it is a Laguerre filter where the tap filters can be replaced by a common pre-filter.

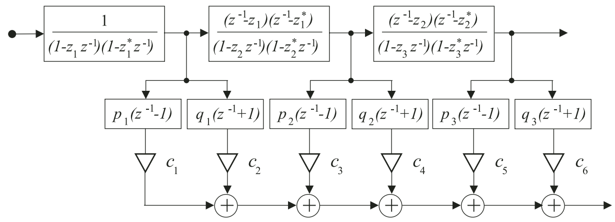

Figure 3: A real Kautz filter [6], where the normalizing coefficients $\{p_i, q_i\}$ can be absorbed into $\{c_i\}$.

In our case, a given target response $h(n)$ is approximated as

$$\hat{h}(n) = \sum_{i=0}^{N} c_i g_i(n), \quad c_i = (h, g_i), \quad (4)$$

that is, as an orthogonal decomposition (or projection) of $h(n)$ with respect to a chosen set of basis functions. Due to the orthogonality, the approximation is inherently in the least-square form, the contribution of each component is explicitly at hand, and the approximation is independent of block (or pole) ordering, which provide useful means for model reduction, monitoring, and extensions. In our case of complex conjugate poles, an equivalent *real Kautz filter* construction [6], depicted in Fig. 3, is used to prevent dealing with complex (internal) signals and filter weights.

Kautz filters and their audio applications are described in more detail in [9]. A small Matlab Toolbox of Kautz modeling can be found in: http://www.acoustics.hut.fi/software/kautz/kautz.htm.

4. MODELING STRING INSTRUMENT SOUNDS

In [4], [5], and [9] we have applied high-resolution signal modeling techniques, FZ-ARMA and Kautz filters, to various cases of string instruments. However, parametric (pole-zero) models that have relatively direct interpretation from a physical point of view and which can be applied in various tasks of sound synthesis and music acoustics, need to be developed. One important aspect is also to keep in mind the perceptual factors, i.e., to include features that are important from a perceptual point of view.

The methods presented here are applicable to sounds where a short excitation, such as a pluck in the guitar or a hit by hammer in the piano, is applied to an approximately linear and time-invariant resonator system[2]. We will first elaborate the acoustic guitar, a popular instrument with well known basic properties, but with surprisingly rich fine details revealed by recent studies [10, 11].

4.1 Case Study 1: The Acoustic Guitar

In the case study our goal is to make an acoustic guitar model where the response of the string, the body, and the excitation are resolved as components of an ARMA model. Although these components are not direct physical properties of the instrument, they can be

[2]Thus for example the violin is not applicable, at least directly.

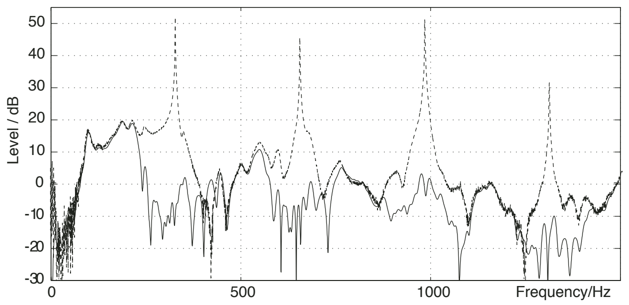

Figure 4: Low-frequency spectrum for the original plucked signal spectrum (dashed line) and the residual spectrum when the FZ-ARMA modeled string modes are removed (solid line).

used in physics-based analysis and synthesis. We will take a typical example of a plucked classical guitar response, recorded in an anechoic chamber[3].

4.1.1 *Extraction of string modes*

The most prominent signal components in a plucked guitar response are the partials of string vibration, see the spectral peaks in Fig. 1. Theoretically each basic mode corresponds to a sinusoid with exponential decay. As already mentioned, the two polarizations of vibration, when summed up in the radiated signal, may introduce beating or two-stage decay in the envelope. Thus two pole pairs are needed if the two-polarization effects are prominent. In practice there may be other phenomena as well [10, 11] that may require even a higher number of pole pairs.

The selection of the number of pole pairs for the FZ-ARMA analysis can be done in several ways. It can be chosen (a) manually by inspection for each partial, (b) it can be iterated, from one pole pair up, until the model fitting does not improve anymore, or (c) it can be a fixed number.

In the following case the sound that we analyzed was played on the open E_4 string (329.6 Hz), plucked in the tirando style. In the spectrum (Fig. 1) there are prominent partials up to about 10 kHz. This range contains 28 partials. When each partial is modeled by two modes (pole pairs), this makes together 112 poles to be estimated by FZ-ARMA analysis. The number of zeros of four per partial was found appropriate. The frequencies of the partials can be approximated easily from the spectrum. Note that due to the inharmonicity of strings, the upper partials are higher in frequency than pure harmonics [3], which is to be taken into account for the frequencies of zooming. In this particular case a zooming factor of $K_{zoom} = 100$ was applied.

The obtained pole set can now be used to evaluate the Kautz filter coefficients c_i (see Eq. (4)), that determine the zeros of the transfer function. When the Kautz model is used for resynthesis to obtain $\hat{h}(n)$, this modeled response is still found to deviate from the original in two ways: the initial transient is different and the body modes of the guitar are not well included. We will next elaborate the separation of the body response.

4.1.2 *Extraction of main body modes*

The string modes can be removed in several ways from the original signal, none of the methods being without problems[4]. We can subtract the Kautz-model response, based on the string poles (see above), from the original, i.e., $h_{resid}(n) = h(n) - \hat{h}(n)$, or we can inverse filter the original by the Kautz-model, i.e., $H_{resid}(z) = H(z)/\hat{H}(z)$. The former one may not fully remove the string mode resonances and the latter one may result in new resonances in the residual if there are sharp zeros in the Kautz model. In the present case we obtained the best results by subtracting from the original signal the FZ-ARMA modeled partials (remapped back from the zoomed frequency domain). Figure 4 shows the spectrum of the residual for the frequency range of 0–1.5 kHz in comparison to the original spectrum.

Now the residual contains the most part of the body resonances and the pluck excitation. Notice that close to the string mode frequencies the residual spectrum cannot be very precise because perfect cancellation of the string modes requires very high accuracy of their estimation, and a part of the body effect may have been included already in the string mode estimation.

In our case study we next apply pole-zero modeling to the residual achieved above to extract a set of body mode poles. This can again be done by various means, but in our case a warped BU-method [12] was used, which combines warping techniques and a particular pole generation procedure for Kautz filtes. The low-frequency region was emphasized by applying a warping parameter value 0.75, which focuses frequency resolution to the lowest body modes. From a generated set of 80 poles two weaker poles were omitted and the remaining set of 39 pole pairs was chosen to represent the resonant body behavior. These body model poles were included in the pool of final Kautz model poles.

4.1.3 *Modeling of source excitation*

When a Kautz model is made using the poles from the string and the body analysis, the initial transient of the reconstruction still deviates notably from the original sound, although the perceptual difference is not very prominent. The effect of plucked transient can be included simply by adding zero-valued poles into the pole set of the Kautz filter. This includes "FIR stages" into the Kautz filter, and if these stages are chosen as the leading block, it actually corresponds to an FIR subfilter. Thus combining the poles from string partials, from the body analysis, and a proper number of zero-valued poles yields an excitation-string-body model, optimal to the pole set in the least-square sense.

4.1.4 *Full Kautz model*

In the E4 string case study we selected 400 FIR block taps to take care of the initial transient. In fact the resulting Kautz model response then replicates the original signal for that number of first samples, i.e., for about 400/44100 $\approx$ 9 ms. Figure 5 depicts the first 40 milliseconds of the original (top) and the resynthesized (bottom) signal. Small differences can be observed in the range of 9 to 15 ms. Figure 6 illustrates the spectral comparison of the original and the resynthesis. The more complex detailed structure found in the original spectrum is partly due to background noise from recording.

In a listening evaluation of the model the most prominent difference of the modeled and the original signal is a minor background noise in the original recording[5]. As another difference, the origi-

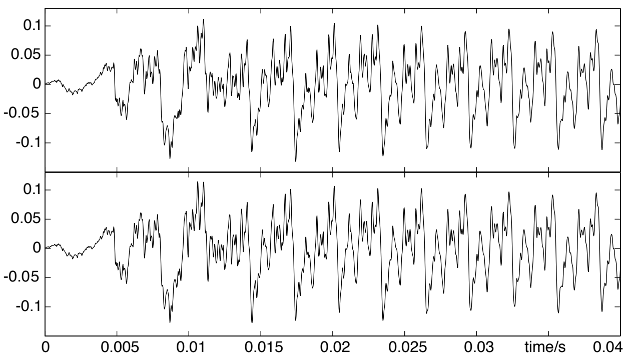

Figure 5: The first 40 ms of the original plucked string signal (top) and the response of the Kautz model (bottom).

[3] We are grateful to Mikael Laurson (guitar player), Henri Penttinen and Jussi Pekonen (recording engineers) for providing the sound examples.

[4] If the body response is available separately, for example from an impact hammer hit, it can be used in body response modeling.

[5] Thus the method works also in denoising of recorded sounds.

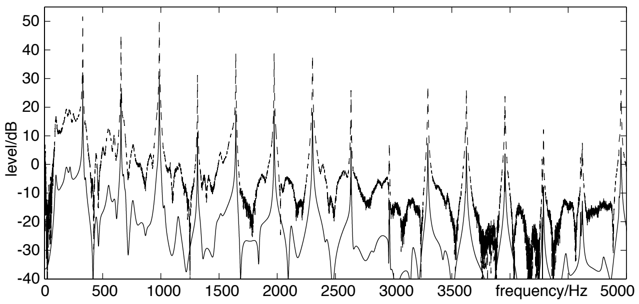

Figure 6: Spectra of the original (upper, dotted) and the resynthesized guitar sound (lower, solid) for 0–5 kHz. The synthesized spectrum is lowered by 20 dB.

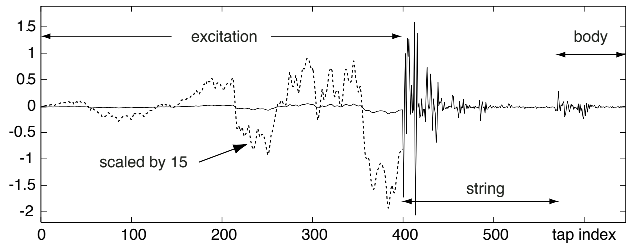

Figure 7: Tap coefficients of the Kautz model for different parts of the model. The FIR part is amplified by factor 15 as the dotted line.

nal one exhibits a bit slower decay of low partials after 1.5 seconds from the beginning (two-stage decay), which is not in the model, but could be included by fine-tuning of the modeling process. Also the body response is perceived by some subjects weaker than in the original one. These differences are hardly noticeable, and by fine-tuning they can be made inaudible. The model order can also be reduced considerably from the case above without radical audible degradation of the result.

It is important to understand the characteristics of the Kautz structure, see Fig. 2. The filter can be seen as a hybrid between series and parallel filter structures. In practice it works more like a parallel filterbank, whereby the substage tag outputs add up to the final output. As a consequence the string partial submodels are almost independent of each other. A single partial can be shifted in frequency or scaled by amplitude, or even eliminated, and the rest of the model remains practically the same. This makes the Kautz model flexible in modifying and morphing string instrument sounds.

This also means that the string part, the body part, and the FIR block work as a parallel connection. This deviates from what would be intuitively desirable, i.e., having them in cascade: excitation $->$ string $->$ body. Particularly the excitation is not fed to the rest of the filter but it is fed to the output, and the rest of the filter is excited by impulse through a delay having the length of the FIR block. Decomposing the model into a cascaded structure remains a challenge for future research.

An interesting view of the model can be obtained by plotting the tap coefficients of the real-valued Kautz model. Figure 7 shows them, separated into sections of the FIR part, the string part, and the body part. The FIR coefficients can be considered as a wavetable, feeding directly to the output. The tap coefficients for the string and body parts determine the amplitudes and phases of the corresponding modes. In Fig. 7 they are ordered from low to high frequencies, and they characterize the amplitudes of the corresponding modes.

5. DISCUSSION AND CONCLUSION

Parametric models of musical instrument sounds can be used for many purposes. Here we include a short discussion on the application scenarios.

- *Sound synthesis*: The models can be used for very high quality sound synthesis, although they have some drawbacks, such as computational load due to high filter order and complexity of controlling the parameters for dynamic sound synthesis. The parameters can also be used as a starting point to calibrate simpler and more efficient synthesis models, such as digital waveguides.
- *Music acoustics*: Although the pole-zero model parameters cannot immediately be interpreted as physical parameters, they can be utilized in many ways in music acoustics studies.
- *Psychoacoustics*: Psychoacoustical studies on musical instrument sounds is an emerging area of research [13]. Parametric models are highly useful there, because realistically modeled sounds can be modified flexibly in order to study how we perceive different attributes and their variation.
- *Parametric audio coding*: The methods presented have potential in parametric audio coding, but they are not as such applicable to polyphonic sounds even when a single instrument such as the guitar is coded. This remains a challenge for future research.

In this study we have shown that high-resolution pole-zero modeling techniques, particularly the FZ-ARMA method and Kautz filters, are powerful ways to decompose plucked and struck string instrument sounds into parametric forms that find applications ranging from sound synthesis to basic research in music acoustics.

6. ACKNOWLEDGMENT

This study is supported by the Academy of Finland projects 53537 and 105629.

REFERENCES

[1] Julius O. Smith, *Principles of Digital Waveguide Models of Musical Instruments*, Chapter 10 in *Applications of Digital Signal Processing to Audio and Acoustics*, (ed. Kahrs and Brandenburg), Kluwer, 1998.

[2] J. Makhoul, 1975. “Linear prediction: A tutorial review,” *Proc. IEEE*, vol. 63 (1975), pp. 561-580.

[3] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, New York, 1991.

[4] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, ”Frequency-Zooming ARMA Modeling of Resonant and Reverberant Systems,” *J. Audio Eng. Soc.*, vol. 50 (2002), no. 12, pp. 1012-1029.

[5] P. A. A. Esquef, M. Karjalainen, and V. Välimäki, ”Frequency-Zooming ARMA Modeling for Analysis of Noisy String Instrument Tones,” *EURASIP J. Applied Signal Processing*, vol. 2003, no. 10, pp. 953-967.

[6] P. W. Broome, “Discrete Orthonormal Sequences”, *J. Assoc. Computing Machinery*, vol. 12 (1965), pp. 151–168.

[7] W. H. Kautz, “Transient Synthesis in the Time Domain”, *IRE Trans. Circuit Theory*, vol. CT–1, pp. 29–39, 1954.

[8] J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*. 2nd Ed., Am. Math. Soc., Rhode Island, 1969.

[9] T. Paatero and M. Karjalainen, ”Kautz Filters and Generalized Frequency Resolution – Theory and Audio Applications,” *J. Audio Eng. Soc.*, vol. 51 (2003), no. 1/2, pp. 27-44.

[10] J. Woodhouse, “On the Synthesis of Guitar Plucks,” *Acta Acustica*, vol. 90 (2004), no. 5, pp. 928–944.

[11] J. Woodhouse, “Plucked Guitar Transients: Comparison of Measurements and Synthesis,” *Acta Acustica United with Acustica*, vol. 90 (2004), no. 5, pp. 945–965.

[12] T. Paatero, “An Audio Motivated Hybrid of Warping and Kautz Filter Techniques,” *Proc. EUSIPCO 2002*, Toulouse, France, Sept. 2002, pp. 627–630.

[13] H. Järveläinen, *Perception of Attributes in Real and Synthetic String Instrument Sounds*. Doctoral thesis. Report no. 68, Helsinki Univ. of Tech., Lab. of Acoustics and Audio Signal Processing, 2003.