Audio Engineering Society

Convention Paper 5526

Presented at the 112th Convention
2002 May 10–13 Munich, Germany

A Multi-Rate Approach to Instrument Body Modeling for Real-Time Sound Synthesis Applications

Balázs Bank[1,2], Giovanni De Poli[2], and László Sujbert[1]

[1] *Budapest University of Technology and Economics, Department of Measurement and Information Systems, Hungary*

[2] *University of Padua, Department of Electronics and Computer Engineering, Italy*

Correspondence should be addressed to Balázs Bank (bank@mit.bme.hu)

ABSTRACT

Real-time physical modeling may have important applications in audio coding and for music industry. Nevertheless, no efficient solutions have been proposed for modeling the radiation of string instrument body, when commuted synthesis cannot be applied. In this novel multi-rate approach, the string signal is split into two frequency bands. The lower is filtered by a long FIR filter running at a considerably lower sampling rate, precisely synthesizing the body impulse response up to 2kHz. In the high frequency band only the overall magnitude response of the body is modeled, using a low-order filter. The filters are calculated from measurements of real instruments. This enables the physical modeling of string instrument tones in real-time with high sound quality.

INTRODUCTION

Physical modeling of musical instruments has been an evolving field in the last two decades. The main feature of this approach is that it concentrates on the sound production mechanism rather then modeling the produced sound itself.

The benefit of the method is that the interaction of the musician can be easily taken into account, since now he/she controls parameters such as plucking force and position instead of partial frequencies and amplitudes. The interaction of the different parts of the instrument, e.g., the coupled vibration of strings, can be also modeled. When a specific note is played on an instrument, its sound affected by the previous notes sounded. This is automatically modeled by the physical approach. By changing the parameters of the model, never-heard sonorities can be created, which still sound natural to the human hear and can be controlled similarly to real instruments.

The physical modeling approach can be also useful for having more insight in how real instruments work. They help to answer such questions as how the physical parameters of the instrument influence the resulted sound. Consequently, they can be used as an aid for instrument design.

As physical models are described by a small number of parameters, they might find applications in audio coding and multimedia. Music coded in such a way would give larger opportunity for interaction to the user. Obviously, physical modeling cannot be the rival of high-quality coding techniques.

A drawback of the approach is the loss of generality and the difficulties in parameter estimation. For example, additive synthesis can be used for any kind of musical instrument sounds by only changing its parameters. On the contrary, here the model structure has to be also changed, especially when the excitation model is concerned. Moreover, different parameter estimation techniques has to be used for different instrument families.

Another disadvantage of physical modeling compared to the signal based methods is the high computational demand. Although computationally efficient methods exist for modeling the effect of the string and the excitation, efficient modeling of the radiation of instrument body could be done by compromising other features of the physical model. In this paper, a new approach is presented for instrument body modeling, which avoids these difficulties.

The paper is organized as follows: first the basic structure of the physical model used in this study is presented. This is followed by the brief description of the digital waveguide modeling, an efficient string modeling technique. The implementation of excitation is also considered. The next section presents the previous techniques proposed for body modeling and shows their difficulties. For avoiding that, a novel multirate method is proposed. Simulation examples are presented in the case of piano, where body modeling seems to be crucial for high quality sound synthesis. A summary concludes the paper.

MODEL STRUCTURE

Since the physical modeling approach simulates the structure of the instrument, the parts of the model correspond to the parts of real instruments. In every string instrument, the heart of the sound production mechanism is the string itself. The string is excited by the excitation mechanism, which corresponds to the hammer strike in the case of the piano, or to the bow in the case of the violin. The string is responsible for the generation of the periodic sound by storing this vibrational energy in its normal modes. One part of this energy dissipates and an other part is radiated to the air by the instrument body. The body can be seen as an impedance transformer between the string and the air, which increases the effectiveness of radiation significantly. The body provides a terminating impedance to the string, therefore it also influences the modal parameters of string vibration, i.e., partial frequencies, amplitudes, and decay times.

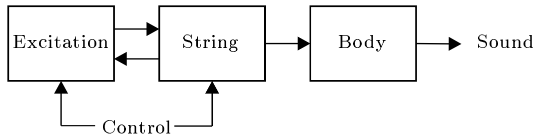

Fig. 1: Model structure.

The model structure is displayed in Fig. 1. It can be seen that the interaction of the string and the excitation is bidirectional. This is because the interaction force depends on the previous state of the string too, and not only on the parameters of the excitation. This is taken into account by introducing a feedback from the string to the excitation model. As mentioned above, the body also influences the body vibration, so the interaction should be bi-directional also in this case. However, in our model the effect of the instrument body is split into two parts. The effect of providing a terminating impedance to the string is taken into account in the string model itself. The body model is responsible for the modeling of radiation properties. This way, the body can be modeled as a straightforward structure.

STRING MODELING

The approach taken here is the digital waveguide modeling [1], which has been found the most effective approach for string modeling so far. The method is based on the time-domain solution of the one-dimensional wave equation. The shape of an infinite and lossless string can be described by the sum of two traveling wave components [2]:

$$y(x,t) = f^+(x-ct) + f^-(x+ct) \tag{1}$$

where x refers to the spatial coordinate, t to time, and y to the transverse displacement of the string. The traveling wave components f^+ and f^- can be any twice differentiable functions.

Time- and spatial-domain discretization of Eq. (1) leads to the representation of two delay lines, one for each traveling wave component. The effects of string losses, dispersion and termination can be taken into account by connecting the delay lines at their boundaries with reflection filters. Thus, the model reduces to a delay line and a digital filter in a feedback loop.

Delay lines can be implemented as circular buffers on DSPs, and filter orders used for the reflection filter are relatively small (2 to 20), thus, this part of the model is well suited for real-time applications.

EXCITATION MODELING

The excitation models are different for the various instrument families. To show some examples, the excitation models of the piano and violin are outlined briefly.

For example, piano hammers can be easily modeled by the discretization of a second-order differential equation [3, 4, 5].

Numerical problems may arise, but they can be solved by various methods [6, 7].

The bow-string interaction of the violin can be modeled by static nonlinearity with varying parameters [8, 9]. Improved models have been also presented [10, 11], but their computational complexity still remains comparable to the one of the digital waveguide.

On the whole, while the excitation models are extremely important for modeling distinctive character of specific instruments, and their quality influences the overall sound significantly, they can be easily implemented in real-time.

BODY MODELING

In most of the cases, the radiation effect of the instrument body is taken into account as a linear filtering operation upon the signal coming from the string model. Thus, the modeling problem reduces to filter design. Unfortunately, the transfer function of real instrument bodies exhibits high modal density, making difficulties for standard filter design algorithms. For high quality sound, high order filters are needed. Their computational complexity can be 10 or 100 times higher than that of needed for the string model. As an example, the pressure-force transfer function of a piano soundboard is shown in Fig. 2. The soundboard was excited by hitting the bridge with an impact hammer. The excitation force and the sound pressure at 2 m distance from the piano were simultaneously recorded. The ratio of their spectra is depicted in this figure. Ideally, the last block of Fig. 1 should have similar transfer function to the one displayed here in Fig. 2.

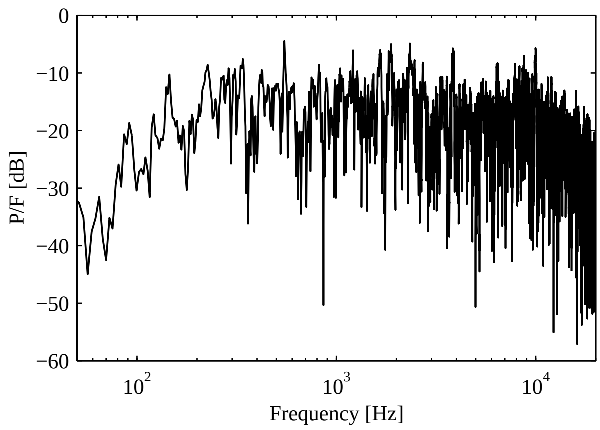

Fig. 2: The transfer function of a piano soundboard

.

Commuted Synthesis

To avoid the problems of filter design and high computational demand, the commuted synthesis technique was presented in [12, 13]. This is based on the idea that if all the elements of Fig. 1 are linear, and the feedback from the string to the excitation is neglected, the system reduces to three linear filters in series. The order of these filters can be commuted, this results in a structure of Fig. 3.

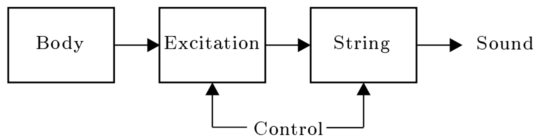

Fig. 3: Commuted synthesis.

As the input signal of the model is a unit pulse, the body filter does not have to be implemented. Its impulse response can be stored in a wavetable whose content is simply fed to the string. The effect of excitation filtering can be also taken into account in the wavetable. In this case, different wavetables have to be stored for the different excitation types for all the notes. This is extremely useful when the goal is the precise resynthesis of the recorded instrument sound, since the output of the model equals with the original up to the length of the excitation table. Astonishing results have been achieved by this method in the case of the acoustic guitar [14, 15].

The drawback of the method is that the excitation model looses its physicality. Now the excitation model is either the part of the wavetable, or implemented as a linear filter. Thus, it is controlled by either switching between wavetables, or changing filter coefficients, rather than varying such a physical parameter as plucking force. The bi-directional interaction of the excitation and the string cannot be implemented anymore, therefore some effects, such as the restrike of a string, cannot be implemented this way. The modeling of certain instruments, e.g., the violin, seems to be impossible even with these restrictions. In some cases, the string model is nonlinear [16], therefore, this method cannot be used. As these limitations seem to be too hard for a general instrument model, we have to find other solutions to the problem.

FIR Filters

When the structure of Fig. 1 is used, the instrument body is implemented as a digital filter. The most straightforward approach to do that is the windowing of the measured impulse response. Hence, it can be implemented as an FIR filter. This technique is simple and capable to provide the best sound quality from a given measurement. On the other hand, high filter orders are required to reach high quality sound.

In the case of the acoustic guitar, it was found that filter orders lower than 1000 do not produce satisfactory sound [17]. For modeling the piano soundboard at a sampling rate of $f_s = 44.1$ kHz, we have found that the sound quality does not increase significantly when the filter length is raised over 2000 tap. Under filter order 1000, the sound starts to loose its character. Consequently, having a filter order between 1000 and 2000 seems to be a reasonable choice. These results coincide with the ones presented in [17].

The computational requirements can be somewhat lessened if the lowest resonances of the instrument body are factored out from the FIR filter and implemented as second-order resonators [17]. In the case of the acoustic guitar, this resulted in 500 tap FIR filters.

Although FIR filters are capable to produce high sound quality, their computational cost is 10 to 100 times higher compared to the cost of the string and excitation models. Nowadays fast DSPs exist, which are able to run such filters in real-time. However, it would be still useful to find other methods with lower computational cost, since they would enable the synthesis of more notes or even different timbres in the same time by the same DSP chip. For multimedia applications, where the system resources are also used for other purposes, it is obvious that having a 1000 tap FIR is too demanding.

IIR Filters

For modeling the violin body, many different filter design techniques were compared in [8]. The final choice was an eighth-order IIR filter designed by minimizing the Hankel norm on Bark scale. Nowadays the implementation of higher-order filters have become possible, but also the quality requirements have increased.

In [17], two IIR filter design methods were compared for modeling the guitar body. It has turned out that IIR filters perform almost the same as FIR filters with the same computational cost. We have found similar results in the case of the piano, i.e., filter orders less than 500 do not produce appropriate results. As also noted in [17], minimum-phase equalization is not a good option, since it destroys the reverberant character of the response. Moreover, in the case of the piano, when experiments were made with minimum-phase filters, it has turned out that they ruin the characteristic attack of the piano and result in an unnatural sound.

Frequency warping can be also used in filter design to give more emphasis to the psychoacoustically more important low frequencies. For the violin model in [8], this approach was used. In [18], warped filter design was proposed for modeling the body response of the acoustic guitar. By this technique, the required filter orders can be reduced significantly. However, either these warped filters need special structures for implementation, or they have to be converted to conventional structures. During conversion, numerical instabilities may arise, when high filter orders are used.

THE NOVEL MULTI-RATE APPROACH

As we are interested in general instrument models, where both the excitation and string models may contain nonlinearity, the commuted synthesis technique cannot be used. Among the filter design approaches, the FIR filter is capable to produce the best sound quality for a given transfer-function measurement. This is because that it preserves not only the overall magnitude response of the instrument body, but also the phase information. Having accurate time-domain response seems to be crucial for the realistic attack of synthesized sounds. Now the problem lies only in how to reduce the computational demand of the FIR filter.

The Structure

Here, a multi-rate approach is presented to avoid the unacceptably high computational cost of the FIR filter, while still maintaining its benefit in preserving the sound characteristics. As shown in Fig. 4, the string signal F_s is split into two frequency bands. The lower is filtered by a long FIR filter $H_l(z)$ running at a considerably lower sampling rate ($f_s' = f_s/8$), precisely synthesizing the body impulse response up to 2 kHz. This means that the same impulse response length (in ms) consumes only the 1/64 part of computation compared to a single-rate filter. This is because the filter length is reduced by a factor of 8 compared to a single-rate FIR filter with the same length in ms, and this shorter filter is run at every 8$^{\text{th}}$ time instant.

In the high frequency band only the overall magnitude response of the body is modeled, using a low-order filter $H_h(z)$ running at the sampling rate of the system ($f_s = 44.1$ kHz). This part of the signal-flow is delayed by N samples to compensate the delay of the downsampling and upsampling operations. The simplification in the high frequency region can be done, because here the human ear has been found to be less sensitive to the temporal evolution of the body response. The crossover frequency of 2 kHz was determined by conducting informal listening tests.

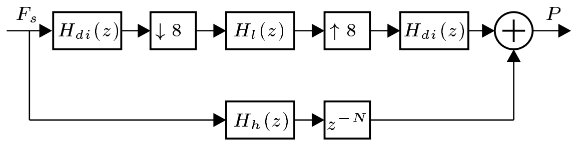

Fig. 4: The multi-rate body model.

For the decimation and interpolation filters, a polyphase FIR filter $H_{di}(z)$ has been used. Note that the interpolation and decimation filters could be different in principle. Here the same filter $H_{di}(z)$ is used for both operation. In a general multi-rate system, large stopband attenuation and small passband ripple would be required, but this would result in long interpolation and decimation filters. However, by taking the advantage of the current situation, having 5 dB passband ripple is an appropriate solution. The passband ripples of the decimation and interpolation filters would be audible, but now their errors can be corrected by changing the magnitude response of the low frequency body filter $H_l(z)$. For stopband attenuation, 60 dB has been found to be sufficient in practice. The filter $H_{di}(z)$ is designed by the `remez` algorithm of MATLAB [19], which produces an equiripple linear-phase filter. The specification is fulfilled by a filter order of 96, which results in 12 operations per cycle when polyphase implementation is considered.

Parameter Estimation

The body filters $H_l(z)$ and $H_h(z)$ can be designed from the measurement of real instruments or by using simulated impulse responses. In this study, measurements were used. The target impulse response $H_t(z)$ is a 2000 tap FIR filter obtained by truncating the measured impulse response. This is lowpass-filtered by a linear-phase FIR filter and then decimated by a factor of 8. This results in a 250 tap FIR filter $\tilde{H}_l(z)$. Now the passband errors of the decimation and interpolation filters have to be corrected. This is computed as follows:

$$H_l(z) = \tilde{H}_l(z)\frac{1}{H_{di}^2(z^{\frac{1}{8}})} \tag{2}$$

where now $H_l(z)$ corresponds to the body filter to be implemented in Fig. 4.

Note that now only the magnitude response of the filter $\tilde{H}_l(z)$ is changed, since the decimation and interpolation filter $H_{di}(z)$ is of linear-phase. However, arbitrary filter types

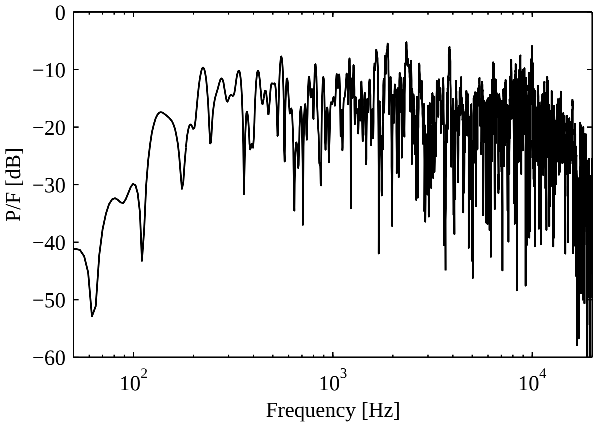

Fig. 5: The mangnitude transfer function of the 2000 tap target filter $H_t(z)$.

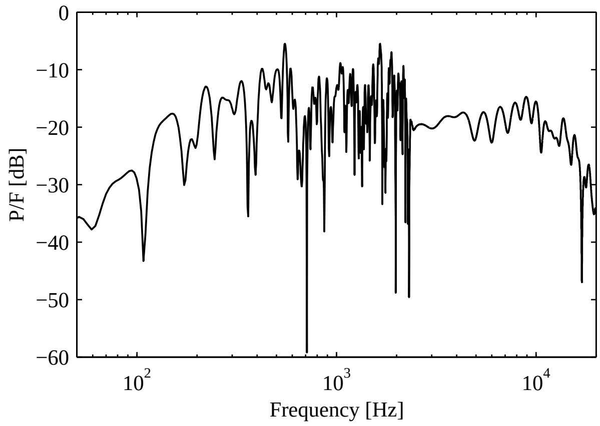

Fig. 6: The magnitude transfer function of the multi-rate body model.

could be used for $H_{di}(z)$, since not only the magnitude, but the phase error of $H_{di}(z)$ can be corrected by Eq. (2).

Now the impulse response of the low frequency chain of Fig. 4 is known. The remaining high frequency part can be easily calculated by subtracting this low frequency response from the target impulse response $H_t(z)$. This way, a 2000 tap FIR filter arise containing energy mainly at frequencies above 2 kHz. This response is then made minimum-phase. This concentrates the energy to the beginning of the impulse response. Then this minimum-phase response is truncated to a length of 50 tap to form the high frequency body filter $H_h(z)$.

Results

As an example, the magnitude response of a piano soundboard model is depicted in Fig. 6. The magnitude response of the target filter $H_t(z)$ is depicted in Fig. 5 for comparison.

It can be seen from the figures that the magnitude response is accurately preserved up to 2 kHz. Although not visible, but so is the phase response. Above 2 kHz, only the overall magnitude response is retained.

The model now consumes 100 operations per cycle, and has a very similar spectral character to the 2000 tap target filter $H_t(z)$. This is because the main features of the instrument body are reproduced. One such feature is the correct overall magnitude response. The other is that it should alter the amplitudes of the consecutive partials to provide roughness in the sound. This is done by the peaks of the transfer function. It can bee seen from Fig. 6, that up to 2 kHz, these resonances equal with the ones of the target response. Above 2 kHz, only small number of resonances are present. However, in the case of high notes, partial frequencies are so much apart, that they fall into different peaks and valleys of the model, thus, their amplitudes are also perturbed.

Another important feature of the instrument body is providing the proper attack of the sound. This is partly fulfilled, since up to 2 kHz the time-domain response of the model equals with that of the target response. The attack of high notes sounds sharper compared to the 2000 tap target filter. This is because the energy of the soundboard response is concentrated to the first 1.1 ms above 2 kHz. Therefore, the attacks of high frequency partials are not smoothed enough by the body filter.

A straightforward solution to the problem is having a higher crossover frequency by decreasing the downsampling factor of the low frequency chain to 4. In this case, the model produces an almost indistinguishable sound compared to the 2000 tap target filter. As a drawback, the computational complexity increases to 190 instructions per cycle.

Other solution could be lengthening the time domain response of the high frequency chain, which corresponds to having sharper resonances in the frequency domain. For doing this, feedback delay networks [20, 21] could be good candidates, since they are capable of producing high modal density at a low computational cost. Implementations with more channels of different sampling rate could be also considered. Combining frequency warping and multi-rate techniques might also be used to reduce the required computational cost or increase the sound quality.

CONCLUSION

Body modeling had been a bottleneck in real-time physical modeling sound synthesis, since it required 10 to 100 times more computation compared to string and excitation models. Here, a novel multi-rate approach was presented for instrument body modeling, which reduces the computational complexity compared to the previous methods by an order of 10 while still maintaining high sound quality. This enables physical modeling in such applications, where computational complexity is an important issue.

The basic idea of this study was born by conducting informal listening tests. From the scientific point of view, it would be interesting to conduct psychoacoustic experiments to find out what are the critical and audible parameters of instrument bodies.

Here, the model parameters were computed from measurements of real instruments. Therefore, the musician is not able to intuitively control the parameters of the model, such as the size or the shape of the soundboard. In most of the cases this is not a limitation, since the body of a real instrument cannot be changed either. On the other hand, when the target impulse response is computed by simulations, e.g., by the finite differences method, it would enable the user to influence the physical parameters of the instrument body in real time. The computationally demanding simulation for computing the target response should be run only at parameter changes. This approach would combine the benefits of both methods and provide the means of creating never heard sonorities.

ACKNOWLEDGEMENT

This work was supported by the EU project MOSART Improving Human Potential and by the Hungarian National Scientific Research Fund OTKA under contract number F035060. The authors are thankful to the Szent István Király Music School for providing the opportunity of piano soundboard measurements.

REFERENCES

[1] Julius O. Smith. Physical modeling using digital waveguides. *Computer Music J.*, 16(4):74–91, Winter 1992.

[2] P. M. Morse. *Vibration and Sound.* McGraw-Hill, 1948. Reprint, (1st ed. 1936).

[3] Antoine Chaigne and Anders Askenfelt. Numerical simulations of piano strings. i. a physical model for a struck string using finite difference methods. *J. Acoust. Soc. Am.*, 95(2):1112–1118, February 1994.

[4] Antoine Chaigne and Anders Askenfelt. Numerical simulations of piano strings. ii. comparisons with measurements and systematic exploration of some hammer-string parameters. *J. Acoust. Soc. Am.*, 95(3):1631–1640, March 1994.

[5] Gianpaolo Borin, Giovanni De Poli, and Auguato Sarti. Sound synthesis by dynamic system interaction. In Denis Baggi, editor, *Readings in Computer Generated Music*, pages 139–160. IEEE Computer Society Press, Los Alamitos, USA, 1992.

[6] Gianpaolo Borin, Giovanni De Poli, and Davide Rocchesso. Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems. *IEEE Trans. on Acoust., Speech, and Sign. Proc.*, 8(5):597–606, September 2000.

[7] Balázs Bank. Nonlinear interaction in the digital waveguide with the application to piano sound synthesis. In *Proc. Int. Computer Music Conf.*, pages 54–58, Berlin, Germany, 2000.

[8] Julius O. Smith. *Techniques for Digital Filter Design and System Identification with Application to the Violin.* PhD thesis, Stanford University, California, USA, June 1983.

[9] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse. On the oscillations of musical instruments. *J. Acoust. Soc. Am.*, 74(5):1325–1345, November 1983.

[10] Stefania Serafin, Christophe Vergez, and Xavier Rodet. Friction and application to real-time physical modeling of a violin. In *Proc. Int. Computer Music Conf.*, pages 216–219, Beijing, China, September 1999.

[11] Tapio Takala, Jarmo Hiipakka, Mikael Laurson, and Vesa Välimäki. An expressive synthesis model for bowed string instruments. In *Proc. Int. Computer Music Conf.*, pages 70–73, Berlin, Germany, 2000. Int. Comp. Music Assoc.

[12] Matti Karjalainen and Vesa Välimäki. Model-based analysis/synthesis of the acoustic guitar. In *Proc. Stockholm Music Acoustics Conf.*, pages 443–447, Stockholm, Sweden, July 1993.

[13] Julius O. Smith. Efficient synthesis of stringed musical instruments. In *Proc. Int. Computer Music Conf.*, pages 64–71, Tokyo, Japan, September 1993.

[14] Vesa Välimäki, Jyri Huopaniemi, Matti Karjalainen, and Zoltán Jánosy. Physical modeling of plucked string instruments with application to real-time sound synthesis. *J. Aud. Eng. Soc.*, 44(5):331–353, May 1996.

[15] Vesa Välimäki and Tero Tolonen. Development and calibration of a guitar synthesizer. *J. Aud. Eng. Soc.*, 46(9):766–778, September 1998.

[16] Tero Tolonen, Cumhur Erkut, Vesa Välimäki, and Matti Karjalainen. Simulation of plucked strings exhibiting tension modulation driving force. In *Proc. Int. Computer Music Conf.*, pages 5–8, 1999.

[17] Matti Karjalainen, Vesa Välimäki, Heikki Räisänen, and Harri Saastamoinen. DSP equalization of electret film pickup for acoustic guitar. In *Proc. 106th AES Conv., Preprint No. 4907*, Munich, Germany, 1999.

[18] Matti Karjalainen and Julius O. Smith. Body modeling techniques for string instrument synthesis. In *Proc. Int. Computer Music Conf.*, pages 232–239, Hong Kong, August 1996.

[19] Matlab 5 manual. 1996. Mathworks, Inc. Natick, Massachusetts, USA.

[20] Jean-Marc Jot and Antoine Chaigne. Digital delay networks for designing artificial reverberators. In *Proc. 90th AES Conv., Preprint No. 3030*, Paris, France, February 1991.

[21] Davide Rocchesso and Julius O. Smith. Circulant and elliptic feedback delay networks for artificial reverbiration. *IEEE Trans. on Speech and Aud. Proc.*, 5(1):51–63, January 1997.