Body Modeling Techniques for String Instrument Synthesis

Matti Karjalainen  
http://www.bat.fi/TMS/Teachers/Karjalainen/Laboratory_of_Acoustics_and_Audio_Signal_Processing.html  
Helsinki University of Technology

Julius Smith  
joestcncs, stanford.edu  
CCLRC, Music Department  
Stanford University

Abstract

Techniques are described for obtaining efficient computational models of stringed instrument resonators such as guitar bodies. Warping the frequency axis to an approximate Bark scale using a first-order conformal map decreases the required body filter order by a factor of 5 to 10 for a given quality level. Structures which implement frequency-warped filters are described. Techniques are described for factoring a body resonator into least-damped and most-damped modes so that the most-damped modes can be computed with the string stored in a shortened look-up table or approximated by a filtered noise burst for computed synthesis.

1 Introduction

All acoustical string instruments have some kind of a body or soundboard which is the main source of sound radiation. Such an "amplification" is necessary since a vibrating string alone has a very limited capability to move air and radiate efficiently. Another function of a body or a soundboard is to add coloration and reverberation to the radiated sound. Since acoustic amplification is more or less based on resonating structures, the spectral content of string output signals is thus changed. Due to a relatively slow decay of the body resonances, the temporal structure of the string signals is also changed to exhibit a reverberant quality.

The third acoustic role of a body or a soundboard is to create complex directivity patterns so that the intensity of the radiated sound field is a function of direction. Combined with room acoustics, this yields spaciousness to the perceived sound.

High-quality real-time sound synthesis of string instruments, based on physical modeling and DSP techniques, has been available for several years (Smith 1983, Karjalainen and Laine 1991, Smith 1993a, Karjalainen et al. 1993). The string itself is very efficiently modeled by digital waveguide filters and extensions of the Karplus-Strong model (Smith 1983, Karplus and Strong 1983, Jaffe and Smith 1983, Smith 1987, Smith 1991, Välimäki et al. 1996), and the excitation may be simply implemented as a wavetable or a set of wavetables. The body (or a soundboard), although in most cases a linear and time-invariant system, is found to be computationally very expensive if a full quality synthesis is desired.

Two kinds of efficient solutions are given in literature. The one consisting of the body response and a simple excitation into a wavetable (Smith 1993a, Karjalainen et al. 1993), is a very practical and straightforward method but lacks parametric control of body features. Body filters with a small or moderate number of resonating modes (e.g., Smith 1983, Kooplick and Massie 1992, 3rd order polynomials) provide another efficient method, but lacking the quality of fullfeatured instrument bodies. These two approaches can be blended to give parametric control over the most important resonances, while retaining the full richness of remaining resonances in theEvt form.

In this paper, we give a survey of body modeling techniques for models-based sound synthesis by covering methods from full-quality fiber models to computed synthesis, as well as hybrid methods, including new body filter designs. Both traditional and new filter structures are utilized, with discussion of estimation techniques for the calibration of model parameters from measured responses of acoustic instruments. The acoustic guitar is used as the primary example in these studies.

2 Body Impulse Responses

The acoustic of string instrument bodies and soundboards is a relatively widely studied topic (Heber and Renssen 1990). A natural starting point for body modeling is to measure or compute the body impulse response. Figure 2a shows the first 100 milliseconds of the impulse response from the body of an acoustic gui.
tar of classical design. The response was measured by tapping the bridge vertically with an impulse hammer (strings were damped) and by measuring the response with a microphone located one meter in front of the sound hole. Figure 1 depicts the magnitude behavior in the frequency domain for the full impulse response.

Figure 1: An example of a body impulse response for an acoustic guitar.

![Impulse response of a guitar body](image)

Figure 2: Magnitude spectrum of the impulse response shown in Figure 1: a) full spectrum, b) low frequencies up to 1 kHz.

![Magnitude spectrum](image)

Figure 3: Time-frequency plot of the guitar body response using short-time Fourier analysis. A Hamming window of 12 ms was used with a 3 ms hop size.

3 Traditional Digital Filters as Body Models

The signal transfer properties from strings to radiated sound can be considered to be linear and time-invariant (LTI) in most string instruments. In this case, an efficient way of implementing the body or soundboard for sound synthesis purposes is by means of digital filtering. Here we first consider the use of traditional filter structures—FIR and IIR filters—as direct implementations of body impulse responses. Then we introduce warped filter techniques and their application to body modeling.

---

Footnote 1: Actually, there is no perfect window size for this analysis since low frequencies require better frequency resolution and high frequencies better temporal resolution. A constant-Q or “wavelet” short-time spectrogram would be closer to an audio transform than the steamed short-time Fourier transform.
FIR and IIR Filters as Body Models

A discrete-time LTI system may be represented using the z transform by

$$H(z) = \frac{B(z)}{A(z)} = \sum_{n=0}^{N} b_n z^{-n} 1 + \sum_{i=1}^{N} a_i z^{-i}$$

(1)

The most straightforward way to realize a known body response is to use the samples of a modeled impulse response, or transfer function, as taps in an IIR filter, for which the coefficients $$a_i$$ in (1) are zero. If $$N$$ is larger than the impulse-response duration, this implements the desired convolution of the input output and the body response yielding a full accuracy body model to the extent that the whole audible portion of the response is available free of noise and artifacts.

As obvious problem with FIR modeling is the filter length $$N$$ and thus the computational expense of the method. In the current guitar example, in order to cover a period of a single decay time constant for the lowest mode, an IIR filter of order $$N = 5000$$ taps is needed when a sampling rate of 22 kHz is used. For a 60 dB dynamic range and full audio bandwidth, an FIR order of about $$N = 25000$$ is required! In practice, using only the first 100 ms of the response is found to be quite satisfactory, which means a 2200 tap filter for a 22 kHz sampling rate. Even this is computationally much more expensive than a model for six guitar strings, and it may be more than a modern signal processing chip can do in real time. The conclusion is that FIR models are generally impractical unless very efficient FIR hardware is available.

The sharply resonating and exponentially decaying components of a body response imply that IIR filters are more appropriate for efficient synthesis models than FIR filters. In order to see how well straightforward all-pole modeling works, we may apply autoregressive (AR) modeling using the autocorrelation method of linear prediction (LP) [Markel and Gray 1976] to the impulse response shown in figure 1. This yields an all-pole filter model where the coefficients $$b_i$$ of (1), for $$i = 2$$ to $$N$$, are equal to zero, are the predictor coefficients, and $$P$$ is the order of the filter. Experiments with all-pole modeling have shown that, in our example, an order of $$P = 500$$ to 1000 is needed to yield a well matched temporal response. Lower filter orders, although relatively good from a spectral point of view, make the lowest resonances decay too fast. This can be addressed using a spectral weighting function (preemphasis) at the expense of the high-frequency fit.

The next generalization with traditional digital filters is to model the impulse response with a pole-zero (or ARMA) model. We have tried this using Prony’s method [Parks and Burrus 1987, pp. 206-209]. The results show that this does not relax the requirements for the order $$P$$, but adding 100 zeros or so to the model improves the fit of the transients’ attack of the impulse response. From the point of view of auditory perception, however, this has only a small effect.

4 Warped Filters

Many filter design and model estimation methods allow for an error weighting function versus frequency in order to control the varying importance of different frequencies. Here, however, we take a different approach: Instead of an explicit weighting function we use frequency scale warping that is in principle applicable to any design or estimation technique. The most popular warping method is to use the bilinear conformal mapping [Churchill 1960, Parks and Burrus 1987] since it is the most general conformal mapping that preserves order. It can be used to warp the impulse response, frequency response, or transfer function polynomials. The warped FFT was introduced by [Oppenheim et al 1971] and warped linear prediction was developed by [Strube 1986]. Generalized methods using the FAM functions have been developed by [Laube et al 1994]. Smith has applied the bilinear mapping in order to design filter models for the violin body [Smith 1983].

The bilinear warping is realized by substituting unit delays by first-order allpass sections, i.e.

$$z^{-1} \rightarrow D_1(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}$$

(2)

This means that the frequency-warped version of a filter can be implemented by such a simple replacement technique. (Modifications are needed to make warped IIR filters realizable.) The transfer function expressions after the substitution may also be expanded to yield equivalent IIR filter of traditional form. It is easy to show that the inverse warping can be achieved with a similar substitution but using -1 instead of $$\lambda$$.

The usefulness of frequency warping in our case comes from the fact that, given a large transfer function $$H(z)$$, we may find a lower order warped filter $$H_w(z)$$ that is a good approximation of $$H(z)$$. For an appropriate value of $$\lambda$$, the bilinear warping can fit the psychoacoustic Bark scale based on the critical band concept, relatively accurately [Strube 1989, Zwicker 1990]. For this purpose, an appropriate formula for the optimum value of $$\lambda$$ as a function of sampling rate is given in [Smith and Abel 1995]. For a sampling rate of 44.1 kHz this yields $$\lambda = 0.7233$$ and for 22 kHz $$\lambda = 0.2258$$. When using the warping techniques, the optimality of $$\lambda$$ in a specific application de-
pends both on salient aspects and the characteristics of the system to be modeled.

Warped FIR (WFIR) Filters

The principle of a warped FIR filter (WFIR) is shown in Figure 4a, which may be written as

\[ B_w(z) = B(z)D^{-1}_{\lambda}(z) = \sum_{i=0}^{M} \beta_i D_i(z) \]  

A more detailed filter structure for implementation is depicted in 4b. As the latter form shows, a warped FIR is actually recursive, i.e., an IIR filter with \( M \) poles at \( z = \lambda \), where \( M \) is the order of the filter.

![Figure 4: Warped FIR modeling: (a) general principle, (b) detailed filter structure for implementation.](image)

A straightforward method to get the tap coefficients \( \beta_i \) for a WFIR filter is to warp the original impulse response and to truncate it by "windowing" the portion that has amplitude above a threshold of interest. (Notice that the bilinear mapping of a signal by (2) is linear but not shift-invariant [Strube 1988].) There exist various formulations for computing a warped version of a signal [Strube 1980, Smith 1983, Laine et al. 1994]. An accurate and numerically stable method is to apply the FIR filter structure of Figure 4a or b with tap coefficients being the amplitudes of the signal to be warped. When an impulse is fed to this filter, the response will be the warped signal. Figure 5 shows the warped (\( \lambda = 0.63 \)) guitar body response as a time-frequency plot for comparison with the original one in Figure 3. As can easily be seen, the warping tends to balance the decay rates and resonance bandwidths for all frequency ranges.

![Figure 5: Time-frequency plot of warped guitar body response. A Hamming window of 54 ms was used with a 3 ms hop size, where * denotes warped time.](image)

Warped IIR (WIR) Filters

When linear prediction is applied to a warped impulse response it yields a warped all-pole filter. Other methods for warped LP analysis (WLP) are studied in [Laine et al. 1994], including an efficient way to compute warped autocorrelation coefficients \( r_w(k) \) directly from the original signal. This is based on the warped delay-line structure of Figure 4a, whereby

\[ r_w(n) = \sum_{\lambda} r_o(\lambda)n \cdot \delta(n) \]  

is summed over a time interval or window of interest. After that, the warped prediction coefficients are achieved from warped autocorrelation coefficients as usual [Markel and Gray 1976] to yield a filter model

\[ H_w(D(z)) = \frac{G_w}{1 + \sum_{\lambda} a_{w}(\lambda)D(z)} \]  

A somewhat surprising observation is that the filter structure of (5) cannot be implemented directly since there will be delay-free loops in the structure for \( \lambda \neq 0 \). (Of course the bilinear mapping, inherent in the filter structure, may be expanded at design time to yield a normal IIR filter. This, however, can lead to numerical difficulties if the filter order exceeds about 20 to 30 in 64-bit double precision floating-point.) Figure 6 depicts two realizable forms of WIR filters. Strube [1986] suggested an approximation in which low-pass sections are used instead of allpass sections (see Figure 6a). In practice, it works only for low orders and warping values \( \lambda \) due to excessive high-frequency attenuation otherwise. The version in Figure 6b is more general for warped pole-zero modeling but has also a more complex structure. The coefficients \( a_{w}(\lambda) \) can be computed from

ICMC Proceedings 1996 235

Karjalainen & Smith
Body Modeling with Warped Filters

For a given quality level, the warped filter strategy using a Bark-scale warping yields a reduction in filter order by a factor 1 to 10. This means that a WBFIR of order M less than 500 gives results similar to those of an FIR filter of order N = 2000. For warped all-pole filters, an order of about R = 100 is equivalent to normal IIR order of about P = 500 to 1000. A body filter resulting from using Froy’s method for a warped filter, with M = 50 to 100 and R = 100 to 200, will represent both transient and decay properties relatively well, although a Bark warping (λ = 0.53, sampling rate 22 kHz) has a tendency to shorten the impulse response of the highest frequencies a bit too much.

The reduction in filter order due to warping translates to a similar reduction in computational complexity only if unit delays and allpass sections are equally complex computationally. In reality, many digital signal processors have hardware support to run ordinary FIR and IIR filters very efficiently. The complexity of the filter becomes more significant than indicated by the order savings. In a typical case, for the TMS20C30 floating-point signal processor, a WBFIR body filter model is only about two times faster than an equivalent normal IIR filter. On the other hand, if the warped filters are converted to conventional structures (using ultra-high-precision computations, e.g., in Mathematica), the extra complexity disappears, since the warping preserves order, it does not have to increase filter complexity except when the number of poles is different from the number of zeros in which case the warping (or unwarping) will introduce new poles or zeros so that their number is the same.

A nice benefit of implementing the body filter in warped form is that λ is available as a qualitative body size parameter. The size parameter can be modulated to obtain new kinds of effects, or it can be used to “morph” among different members of an instrument family.

5 Body-Model Factoring

Figure 7: Schematic model of a stringed instrument in which the string and resonator are connected relative to their natural ordering.

Commutated synthesis is a technique in which the body resonator is commutated with the string model, as shown in Fig. 7, in order to avoid having to implement a large body filter at all [Smith 1995a, Karjalainen et al. 1995]. In commutated synthesis, the excitation (e.g., plucking force versus time) can be convolved with the resonator impulse response to provide a single aggregate excitation signal. This signal is short enough to store in a look-up table, and a note is played by simply summing it into the string.

A valuable way of shortening the excitation table is to commutate synthesis in factor the body resonator into its most-damped and least-damped modes. The least-damped modes are then commutated and combined with the excitation in impulse-response form. The least-damped modes can be left in parametric form as recursive digital filter sections. Advantages of this factoring include the following:

- The excitation table is shortened.
- The excitation table signal-to-quantization noise ratio is improved.
- The most important resonances remain parametric, facilitating real-time controls.
- Multiple body outputs become available.
- Resonators may be already available in a separate effects unit, making them “free.”
- A memory vs. computation trade-off is available for cost optimization.

Mode Extraction Techniques

The goal of resonator factoring is to identify and remove the least-damped resonant modes of the...
impulse response. In principle, this means se-
certaining the precise resonance frequencies and
bandwidths associated with each of the narrow-
est "peaks" in the resonator frequency response,
and dividing them out via inverse filtering, so
they can be implemented separately as resonators
in cascade. If in addition the amplitude and
phase of a resonance peak are accurately meas-
urable in the complex frequency response, the
mode can be removed by complex spectral sub-
traction (equivalent to subtracting the impulse-
response of the resonant mode from the total
impulse response); in this case, the parametric
modes are implemented in a parallel bank as in
[Bradley and Stonick 1995]. However, in the par-
allel case, the residual impulse response is not
readily commuted with the string.

Various methods are available for estimating the
mode parameters for inverse filtering:

- Amplitude response peak measurements
- Weighted digital filter design
- Linear prediction
- Sinusoidal modeling
- Line impulse-response analysis

In the body factoring example presented be-
low, the frequency and bandwidth of the main
Helmholtz air mode were measured manually us-
ing an interactive spectrum analysis tool. It is also
easy to automate peak-finding in FFT magnitude
data, as is routinely done in sinusoidal modeling,
discussed further below.

![Figure 8: Illustration of one way to determine the parameters of a least-damped resonant mode.](image)

Many methods for digital filter design support
spectral weighting functions that can be used to
focus in on the least-damped modes in the fre-
quency response. One is the weighted equation-
error method which is available in the matlab
invfreq() function. Figure 8 illustrates use of
it in a simple synthetic example with only one
frequency response peak in the presence of noise.
Unless the weighting function is very tight around
the peak, its bandwidth tends to be overestimated.

Another well known method for converting the
least-damped modes into parametric form is Lin-
ear Prediction (LP) followed by polynomial fac-
torization to obtain resonator poles. LP is par-
specially good at fixing spectral peaks due to
the nature of the error criterion it minimizes [Smith
1983, pp. 43-56]. The poles closest to the
unit circle in the z plane can be chosen as the least-
damped part of the resonator. It is not known
that when using LP to model spectral peaks for
extraction, orders substantially higher than twice
the number of spectral peaks should be used.

Another way to find the least-damped mode pa-
rameters is by means of an analytical model of the
body impulse response such as is often used to de-
termine the string loop filter in waveguide string
models [Smith 1993b, Yliniemi et al. 1996]. (See,
for example, Serra and Smith 1990 for further details on
sinusoidal modeling and supporting C software).
The sinusoidal amplitude envelopes yield a par-
ticularly robust measurement of bandwidth. The-
oretically, the modal decay should be exponen-
tial. Therefore, on a dB scale, the amplitude en-
velope should decay linearly. Linear regression
can be used to fit a straight line to the measured
log-amplitude envelope of the impulse response of
each long-cringing mode. Note that even when am-
plitude modulation is present due to mode cou-
plings, the ripples tend to average out in the re-
gression and have little effect on the slopes mea-
asurement. This robustness can be enhanced by
starting and ending the linear regression on local
maxima in the amplitude envelope.

All methods useful with inverse filtering can be
modified based on the observation that late
in the impulse response, the damped modes have
died away, and the least-damped modes dominate.
Therefore, by discarding initial impulse-response
data, the problem in some sense becomes "easier" at
the price of working closer to the noise floor.

High-Frequency Modes ≈ Noise

Figure 9b suggests that the many damped modes
remaining in the shortened body impulse response
may not be clearly audible as resonances since
their decay time is so short. This is confirmed by
listening to shortened and spectrally fattened
body responses which sound somewhere between
a click and a noise burst. These observations sug-
gest that the shortened, flattened body response
can be replaced by a psychoacoustically equiva-
tent noise burst, as suggested for modeling piano
soundboards [Van Dyne and Smith 1995]. Thus,
The signal of Fig. 9b can be synthesized qualitatively by a white noise generator multiplied by an amplitude envelope. In this technique, it is helpful to use LP at the residual signal to flatten it. As a refinement, the noise burst can be time-varying filtered so that high frequencies decay faster [Van Dyne and Smith 1995]. Thus, the stringed instrument model may consist of:

- noise generator
- excitation amplitude-shaping
- time-varying lossy lowpass
- string model
- parametric resonators
- multiple outputs.

In addition, properties of the physical excitation may be incorporated, such as comb filtering to obtain a virtual pick or hammer position. Multiple outputs provide for individual handling of the most important resonant modes and facilitate simulation of directional radiation characteristics [Vanhasselt et al. 1996].

**Body Factoring Example**

![Image](https://example.com/image1)

**Figure 9:** Impulse response of a classical guitar body before and after removing the first peak (main air resonance) via the inverse filter defined by Eq. (6), with $a_1 = -1.9953$ and $a_2 = 3.9972$.

Figure 9a shows the measured guitar body impulse-response data plotted in Fig. 1 but extended to its full duration. Figure 9b shows the same impulse response after factoring out a single resonant mode near 100 Hz (the main Helmholtz air mode). As can be seen, the residual response is considerably shorter than the original.

Figure 10a shows the measured guitar body impulse-response data plotted in Fig. 1 but extended to its full duration. Figure 10b shows the same impulse response after factoring out a single resonant mode near 100 Hz (the main Helmholtz air mode). As can be seen, the residual response is considerably shorter than the original.

The modal bandwidth used in the inverse filtering was chosen to be 10 Hz which corresponds to a Q of 46 for the main air mode. If the bark-warping is done using a first-order conformal map [Smith and Abel 1995], its inverse pre-warp filters order [Smith 1983, pp. 61-67]. Applying the inverse warping to the parametric resonator drives its pole radius from 0.99858 in the Bark-warped z-plane to 0.99997 in the unwarped z-plane.

Note that if the inverse filter consists only of two zeros determined by the spectral peak parameters, other portions of the spectrum will be modified by the inverse filtering, especially at the next higher resonance, and in the linear trend of the overall spectral shape. To obtain a more localized mode extraction (useful when the procedure is to be repeated), we define the inverse filter as

$$H(z) = \frac{A(z)}{A(zr)} \approx \frac{1 + a_1z^{-1} + a_2z^{-2}}{1 + a_1r^{-2} + a_2r^{-4}}$$

where $A(z)$ is the inverse filter determined by the peak frequency and bandwidth, and $A(z/r)$ is the same polynomial with its roots contracted by the factor $r$. If $r$ is close to but less than 1, the poles and zeros substantially cancel far away from the removed modal frequency so that the inverse filter has only a local effect on the frequency response. In the computed examples, $r$ was arbitrarily set to 0.9, but it is not critical.

6 Conclusions

A number of techniques were discussed for improving the quality of virtual stringed instrument resonators and reducing implementation cost. These methods make it possible to simulate high quality stringed instruments using inexpensive software algorithms. As an example, a basic classical guitar model requires less than two percent of a 120
MRX Pentium processor for each actively vibrat-
ing string. Therefore, such models are practical
today, even for low-cost multimedia applications.
For the future, physical models suggest how to
make good use of increased processing power as it
comes along.

References
[Bradley and Stonick 1992] Bradley, K., and V. Ston-
ick. 1992. "Autoquantized Analysis and Compu-
tationally Efficient Synthesis of Acoustic Guitar
Strings and Bells." In: Proc. IEEE Workshop on
[Fletcher and Rossing 1990] Fletcher, N. M., and T. D.
Rossing. 1990. The Physics of Musical Inst-
Guitar on a Floating-Point Signal Pro-
Acoustics, Speech, and Signal Processing, Toronto,
Synthesis of Guitar on a Floating-Point Signal Pro-
Acoustics, Speech, and Signal Processing, Toronto,
[Kaajalainen et al. 1993] Kaajalainen, M., V. Viilimäki,
Music Comp., Tokyo. Computer Music
Association.
1993. "Acoustic and Psychoacoustic Simulation of
Music Comp., Tokyo. Computer Music
Association.
[Laine et al. 1994] Laine, U. K., M. Kaajalainen,
and T. Alho. 1994. "Linear Prediction (WLP)
Acoustics, Speech, and Signal Processing,
Atlanta, Georgia, pp. III:265-302.
Gray. 1979. Linear Prediction of Speech. New York:
Springer-Verlag.
Johnson, and K. Steiglitz. 1971. "Computation of
Spectra with Unequal Resolution Using the Fast
[Parks and Burrus 1977] Parks, T. W., and C. S.
Wiley and Sons, Inc.
1960. "A Spectral Modeling Synthesis System Based on a Determin-
itic plus Stochastic Decomposition." Computer
http://www.nua.upf.es/ serra/.
Digital Filter Design and System Identification with
Dept., Stanford University.
Association.
using Digital Waveguides." Computer Music
J., 16(1):74-91. Special issue: Physical Modeling of
Musical Instruments, Part I.
of Stringed Musical Instruments." Pages 61-71 of:
Association.
[Smith 1993b] Smith, J. O. 1993b. "Structured Sam-
pling Synthesis: Automated Construction of Phys-
ical Modeling Synthesis Parameters and Tables from
Recorded Sounds." (Poster presentation outline, 51
pages), ICMA Associates Conference, May.
1992. "The Bark Bilingual Transformer." In:
Proc. IEEE Workshop on Appl. Signal Processing
crma.stanford.edu/.
of America, 68(4):1071-1076.
[Viilimaki et al. 1993] Viilimaki, V. J., V. Kuopio-
Application to Real-Time Sound Synthesis." J. Audio
[Van Dyne and Smith 1966] Van Dyne, S. A., and J.
O. Smith. 1966. "Developments for the Com-
Computer Music Conf., Banff Computer Music
Association.
New York: Springer-Verlag.