MODELING PLUCKED GUITAR TONES VIA JOINT SOURCE-FILTER ESTIMATION

Raymond V. Migneco and Youngmoo E. Kim

Department of Electrical and Computer Engineering,
Drexel University Philadelphia, U. S. A.
{rmigneco, ykim}@drexel.edu

ABSTRACT

Physical models for plucked string instruments can produce high-quality tones using a computationally efficient implementation, but the estimation of model parameters through the analysis of audio remains problematic. Moreover, an accurate representation of the expressive aspects of a performance requires a separation of the performer's articulation (source) from the instrument's response (filter). This paper explores a physically-inspired signal model for plucked guitar sounds that facilitates the estimation of both string excitation and resonance parameters simultaneously. We present the application of the joint source-filter model in an analysis-synthesis framework to plucked-guitar recordings, and we demonstrate that our system is particularly adept at capturing and replicating the characteristic sounds resulting from various plucking styles. By explicitly modeling string articulations, we believe this system provides insight towards capturing the expressive intentions of a performer from the audio signal alone.

***Index Terms*—** physical modeling, convex optimization, musical instrument synthesis

1. INTRODUCTION

For the past two decades, physical and physically-inspired modeling systems have emerged as a popular method for simulating plucked-string instruments since they are capable of producing high-quality tones with computationally efficient implementations. The first of such modeling systems was the Karplus-Strong Algorithm, which permitted synthesis of plucked-string timbres with low computational cost compared to FM or additive synthesizers. This algorithm works by simply circularly shifting the contents of a randomly initialized delay line through a low pass filter, which produces a tone whose pitch is determined by the delay line's length and the sampling frequency [1].

Smith facilitated the development of digital waveguide models, which provide a clear semblance to the physical phenomena involved with exciting an instrument. For plucked-strings, waveguide models are based on discretizing the solution for a lossy, vibrating string, which consists of two disturbances traveling in opposite directions from the string's initial displacement, into two delay lines [2]. Thus, the delay lines are initialized with the string's initial shape and its motion is simulated by circularly shifting the samples through the delay lines and applying the appropriate boundary conditions at the termination points.

Despite their attractiveness, calibration of physically inspired models relies on offline analysis to determine the optimal parameters for accurate re-synthesis. Moreover, in most implementations expressive control of synthesized sounds consists of manipulating stored excitation signals to adjust their timbral parameters, without considering the player-string interaction during the analysis.

In this paper, we propose a method to jointly estimate the source and filter parameters of a physically-inspired signal model of electric guitar strings. By employing a joint estimation approach, we seek to simultaneously capture variations in a player's "pluck" (source) and string response behavior (resonant filter), which combine to affect instrument timbre, resulting in a musically expressive performance. Such a system has several applications including the transcription of guitar performance with expressive cues and synthesizing high-quality guitar tones with more intuitive expressive control.

2. SDL MODEL FOR GUITAR SYNTHESIS

While waveguides provide a physically clear model for plucked-guitar simulation, they are not adept for certain tasks, such as calibrating model parameters from a recorded performance. This is due to waveguide synthesis depending on the string's initial conditions, including amplitude and plucking position, which are typically not available from a recording. Instead, many plucked-guitar synthesis systems utilize the so-called *single delay-loop* (SDL) model for analysis-synthesis tasks [3], since they permit guitar analysis-synthesis tasks from a source-filter perspective. This section will outline the basic SDL structure and discuss existing calibration and synthesis techniques using the model.

This work is supported by NSF Award IIS-0644151.

2.1. Single delay-loop components

The basic single delay-loop (SDL) structure is shown in Figure 1, and is derived by consolidating the delay lines from the bidirectional waveguide model into a single delay and omitting the string's initial conditions [2, 4]. To synthesize

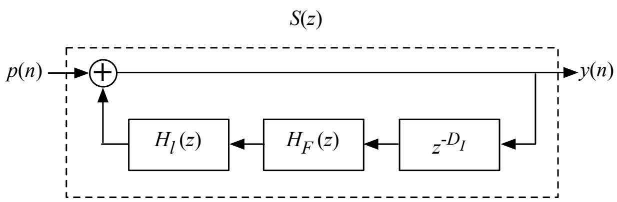

Fig. 1. Single delay-loop (SDL) model for plucked-string synthesis

a plucked-string tone with a desired pitch, f_0, the total delay, D, in the feedback loop of Figure 1 should be set such that

$$D = \frac{f_s}{f_0} \tag{1}$$

where f_s is the sampling rate. Since D is typically a non-integer, the bulk delay line, z^{-D_I}, implements the integer component of D while the remainder is handled by a fractional delay filter, $H_F(z)$. All-pass and Lagrange interpolation filters are typically used to implement the fractional delay instrument synthesis systems [5].

$H_l(z)$ is referred to as the *loop filter*, which simulates the frequency-dependent decay characteristics of the string. $H_l(z)$ can be viewed as a consolidation of all spatial losses incurred by the disturbances propagating through the waveguide model. The loop filter used for this task is typically a first order, IIR filter shown in Equation 2 since it has few parameters to calibrate [6].

$$H_l(z) = \frac{g\,(1-\alpha_0)}{1-\alpha_0 z^{-1}} \tag{2}$$

The gain of $H_l(z)$ is restricted to a maximum gain of unity such that $0 < g \leq 1$ and the pole must be in the range $0 < \alpha_0 < 1$ for stability.

An important distinction between the SDL and the waveguide models is that the SDL model treats plucked-string synthesis as a convolution between an excitation signal, $p(n)$, and a string filter, $S(z)$, rather than a direct simulation of waves propagating along a string. The advantage of the SDL, however, is that it facilitates analysis-synthesis tasks without requiring knowledge of the string's exact physical configuration. The task at hand is determining the proper excitation signal and filter parameters for the model.

2.2. SDL calibration

Loop filter calibration is a primary topic in much of the prior work on plucked-guitar synthesis using the SDL model, since it models the string's natural decay characteristics [4, 6, 7, 8]. Calibration systems typically employ short-time Fourier analysis on a recorded tone to extract the magnitude trajectories of the first 10-20 partials. These trajectories are used to derive loop filter parameters in accordance with Equation 2 that satisfy a least-squares criterion [6, 8]. Laurson et al. proposed an iterative calibration scheme that searches for an optimal set of loop filter parameters, which yield a synthetic signal that best matches the amplitude envelope of a recorded tone [7, 9].

2.3. Plucked-string re-synthesis

Since the SDL model treats guitar synthesis as a source-filter convolution, the excitation signal, $p(n)$, is computed by inverse filtering $y(n)$ in the frequency domain with $S(z)$ to yield

$$\begin{aligned} P(z) &= Y(z)S^{-1}(z) \\ &= Y(z)\frac{1-\alpha_0 z^{-1} + g(1-\alpha_0)H_F(z)z^{-D_I}}{1-\alpha_0 z^{-1}} \end{aligned} \tag{3}$$

where $S^{-1}(z)$ is obtained by inverting the transfer function of the feedback loop in Figure 1. Employing this technique on recordings of acoustic instruments is referred to as *commuted synthesis*, since $p(n)$ is an aggregate signal that captures the excitation and resonant body effects without explicitly modeling the acoustic body [10]. In contrast to acoustic guitars, unprocessed electric guitar tones have less pronounced resonant effects and we are not interested in modeling them. However, commuted synthesis is still useful since it captures the perceptual and physically relevant aspects of the excitation.

3. MODELING GUITAR STRING EXCITATION

While the calibration and synthesis methods presented in Sections 2.2 - 2.3 yield high-quality plucked-guitar timbres, far less research has emphasized deriving parameters from the excitation signal that indicate how the guitarist articulates the string. In this section, we briefly overview existing techniques for expressive guitar modeling and propose a model for the SDL excitation signal that captures expressive intentions by the guitarist.

3.1. Background

Cuzzucoli et al. proposed synthesizing classical guitar performance by studying the finger-string interaction incurred with different plucking styles [11]. This analysis involved studying the string behavior resulting from how quickly the string was displaced by the guitarist's finger. It was shown that this behavior could be incorporated into the bidirectional waveguide model, but no methods were provided for determining the parameters from a recording.

Erkut and Laurson incorporated expressive guitar synthesis into the SDL model by manipulating a reference excitation signal to achieve a desired level of musical dynamics (relative "loudness"). This approach involves using the log-magnitude difference between the reference and desired excitation to design a "pluck-shaping" filter that yields the desired dynamic level [7]. A limitation of this method, however, is it assumes that the string filter does not vary with excitation, which is contrary to actual guitar performance.

3.2. Parameterizing the SDL excitation

Rather than limiting expressive analysis and synthesis to manipulating a reference excitation signal, we feel that accurately modeling the guitar's perceptual and expressive characteristics requires simultaneously accounting for variation of source and filter parameters in the SDL model. This approach, however, requires a model for the excitation signal, $p(n)$, that can encapsulate many different string articulations by the guitarist.

To determine if $p(n)$ could be parametrically represented, we extracted the excitation signals corresponding to recordings of various plucking styles using the following approach:

1. Vary the relative plucking strength (medium and soft)
2. Alter the plucking mechanism (pick or thumb)
3. Calibrate the string filter, $S(z)$, using the technique proposed by Välimäki [6]
4. Extract $p(n)$ through inverse filtering using Equation 3

The results of this analysis for four different plucks on the open 6th string of an electric guitar are shown in Figure 2.

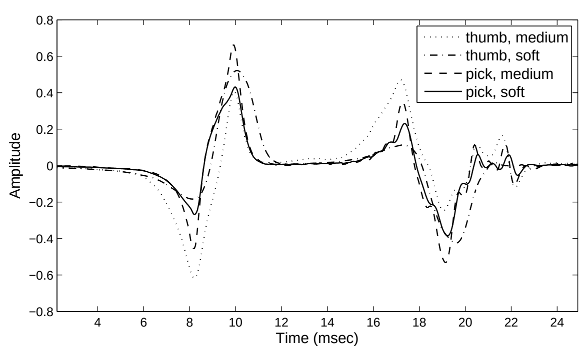

Fig. 2. Excitation signals obtained from various picking styles on the 6th string of an electric guitar (E_2, $f_0 = 82.41$ Hz).

We observe that each excitation signal in Figure 2 exhibits a common contour pattern, which has a physical relation to the bidirectional waveguide model. The regions where signals have maximum amplitude displacement correspond to time instances where the traveling waves constructively interfere, while regions with minimum displacement are due to destructive interference. We also observe that the excitation signals differ in terms of their amplitude and contour smoothness, which can be attributed to the effects of the plucking strength and mechanism on the string's initial shape.

Based on our observations from Figure 2 we propose modeling the SDL excitation signal as a piece-wise polynomial function where the "pieces" are adjoined near regions of maximum and minimum displacement. This function has the form

$$p(n) = \sum_{j=1}^{J} \sum_{k=0}^{K} c_{j,k} n^k \tag{4}$$

where J indicates the number of segments used to partition the excitation and K indicates the polynomial order used to model each segment.

4. JOINT SOURCE-FILTER ESTIMATION

Using the excitation model presented in Section 3.2, we present a framework for jointly estimating the source and filter parameters directly from a recorded signal. Our approach involves finding the parameters that minimize a modeling error using convex optimization.

4.1. Error minimization

To estimate the optimal SDL source and filter parameters, we wish to minimize the model error

$$e(n) = p(n) - \hat{p}(n) \tag{5}$$

where $p(n)$ is the "true" excitation model from Equation 4 and $\hat{p}(n)$ is the residual obtained by inverse filtering. However, as shown in Equation 3, $S^{-1}(z)$ contains a pole at $z = \alpha_0$, which does not allow us to compute $\hat{p}(n)$ independently. Rather, the pole in $S^{-1}(z)$ places dependencies on Equation 5, thus breaking the convexity of the problem.

To overcome this limitation, we remove the pole from $S^{-1}(z)$ (or conversely the zero in $S(z)$) so $\hat{p}(n)$ can be obtained by FIR filtering. This simplification ignores the phase contribution from the pole, though we expect it to have a negligible effect on the magnitude response of $S^{-1}(z)$ since the values obtained for α_0 are typically close to the unit circle [6]. With this simplification, we can express the model error in the frequency domain by taking the z-transform of Equation 5 and substituting for $S^{-1}(z)$, which yields:

$$E(z) = P(z) - Y(z)\left(1 - \alpha_0 z^{-1} - g(1-\alpha_0)H_F(z)z^{-D_I}\right) \tag{6}$$

Equation 6 can be fully expanded by choosing the desired order, N, of the fractional delay filter, $H_F(z)$, which we assume is an FIR filter with the following form:

$$H_F(z) = \sum_{n=0}^{N} h_n(n) z^{-n} \tag{7}$$

$H_F(z)$ introduces additional group delay depending on the order N, which is computed as $\lfloor N/2 \rfloor$ [5]. To avoid detuning the SDL, we compensate for this extra delay by subtracting it off the bulk delay line, z^{-D_I} such that $D_I^* = D_I - \lfloor N/2 \rfloor$.

We substitute Equation 7 and the modified delay term, D_I^*, into Equation 6 and take the inverse z-transform to obtain

$$e(n) = p(n) + \alpha_0 y(n-1) + \beta_0 y(n - D_I^*) + \\ \beta_1 y(n - D_I^* - 1) + \cdots + \beta_N y(n - D_I^* - N) - y(n) \tag{8}$$

where $\beta_k = g(1-\alpha_0)h_k$ for $k = 0, 1, \ldots, N$. For a more compact representation, we can express Equation 8 in matrix form

$$\mathbf{e} = \mathbf{H}\mathbf{x} - \mathbf{y} \tag{9}$$

where $\mathbf{H}$ is a matrix containing the time indices for the excitation model and the time-shifted samples of $y(n)$ and $\mathbf{x} = \begin{bmatrix} c_{1,0} & \cdots & c_{1,K} & c_{J,0} & \cdots & c_{J,K} & \alpha_0 & \beta_0 & \cdots & \beta_N \end{bmatrix}^T$ is a column vector containing the source-filter coefficients we wish to estimate. From Equation 9, we see that the parameters for $p(n)$, which include the number of segments and their locations, must be fully specified so that $\mathbf{H}$ is properly aligned with $\mathbf{y}$. This requires determining the boundary locations and the desired polynomial order for the segments in $p(n)$.

4.2. Solving the convex optimization

We solve for the unknown source-filter coefficients using quadratic programming by taking the L_2 norm of Equation 9 [12].

$$\min_{\mathbf{x}} \|\mathbf{e}\|^2 = \min_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 \tag{10}$$

$$\text{Subject to } 0 < \alpha_0 < 1 \text{ and } \beta_k = g(1-\alpha_0)h_k$$

A caveat of Equation 10 is that the β_k coefficients depend on two unknowns, the gain and pole of the loop filter. Quadratic programming packages support dependencies between two unknowns, but the problem is not convex if a term depends on multiple unknowns. In practice, we handle this by varying g in small increments in the range $(0, 1]$ and solving Equation 10 for each value of g. The optimal solution is chosen by determining the set of source-filter parameters, $\mathbf{x}$, that minimizes the error between the original and re-synthesized signal.

5. IMPLEMENTATION

In this section, we present our system for implementing the joint-source filter estimation scheme for expressive guitar modeling. A diagram of the proposed system is shown in Figure 3 which outlines the major subtasks required for estimating the model parameters and initializing the optimization problem.

5.1. SDL tuning

In order to properly tune the SDL, we must determine the pitch of $y(n)$. We employ the well-known short-time autocorrelation function to estimate the pitch from each frame for 2-3 seconds after the initial "attack" of the tone [13]. The average pitch is computed from all analyzed frames and used to determine the total delay, D, in the SDL using Equation 1.

We implement the fractional delay filter using 5th order, FIR Lagrange interpolation filters, which are commonly used in SDL-based synthesis systems [3, 6, 8]. The impulse response coefficients of these filters are relatively easy to compute and they yield flat magnitude and group delay responses over a wide bandwidth [5].

5.2. Segment estimation

Minimizing the model error in Equation 9 requires determining the boundaries of the segments comprising the excitation model. This task is not straight-forward, but by recognizing that the excitation is in phase with the first period of $y(n)$ (see Figure 5), we can use it as a reference for determining the logical boundaries of $p(n)$.

Using the estimated pitch value, we isolate the first period of $y(n)$ and identify the samples corresponding to zero crossings in the signal and its derivative, which yields inflection points. However, this method is sensitive to spurious peaks from noise, which we do not want to explicitly model. To overcome this, we employ a heuristic method where a "strength" is calculated for each zero crossing in $y(n)$ and its derivative, which is based on amplitude differences between nearby local minimum and maximum peaks. Empirically, we find that retaining 7-10 of the "strongest" samples sufficiently identifies the main contours for $p(n)$ and selecting $K = 5$ as the polynomial order accurately models each contour's shape.

5.3. Iterative optimization

We employ an iterative approach to identify the optimal source-filter parameter set, as discussed in Section 4.2. We vary g in small increments and solve Equation 10 for each value. The mean-squared error (MSE) between the original and synthesized signals is computed for each solution of Equation 10 and the optimal parameter set is chosen as the one yielding the minimum error.

The MSE function for plucks acting on the guitar's 3rd string is shown in Figure 4. This plot demonstrates that the MSE converges to a minimum for each pluck, and the corresponding gain values are clustered, thus demonstrating a global property of the string filter.

6. PRELIMINARY RESULTS

We evaluated our system's ability to jointly estimate the source and filter parameters by analyzing recorded perfor-

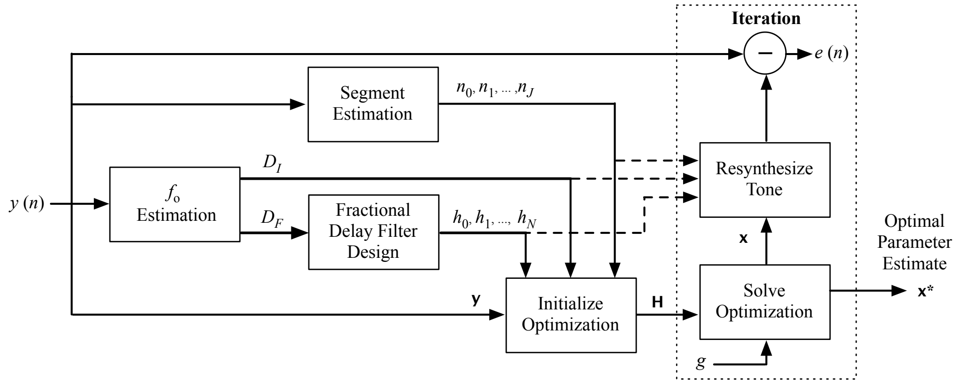

Fig. 3. System diagram for joint source-filter estimation of plucked guitar tones.

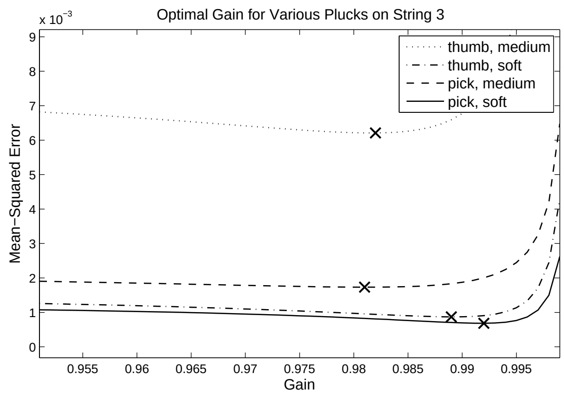

Fig. 4. Mean-squared error (MSE) as a function of g for various plucked tones on the 3rd string. The optimal parameter set belongs to g which minimizes the MSE, which are marked for each pluck.

mances of a guitarist exciting each of the 6 open strings on an electric guitar. We generated 4 recordings per string, each of which features the guitarist varying the plucking strength (medium vs. soft) and plucking mechanism (pick vs. thumb). We implement the subtasks in the parameter estimation as described in Section 5 using the system shown in Figure 3.

Figure 5 demonstrates the application of our system to a recorded signal. We observe that our system is adept at approximating the boundaries of the excitation signal in the first period of the recorded tone, which permits accurate modeling of the excitation signal. Through informal listening tests, we observe that the synthesized signals preserve the perceptually important characteristics of the original tones, including the timbral characteristics resulting from the guitarist's plucking style [1].

Table 1 summarizes the mean-squared error computed between the original and resynthesized plucks under analysis. We observe that the average modeling error generally

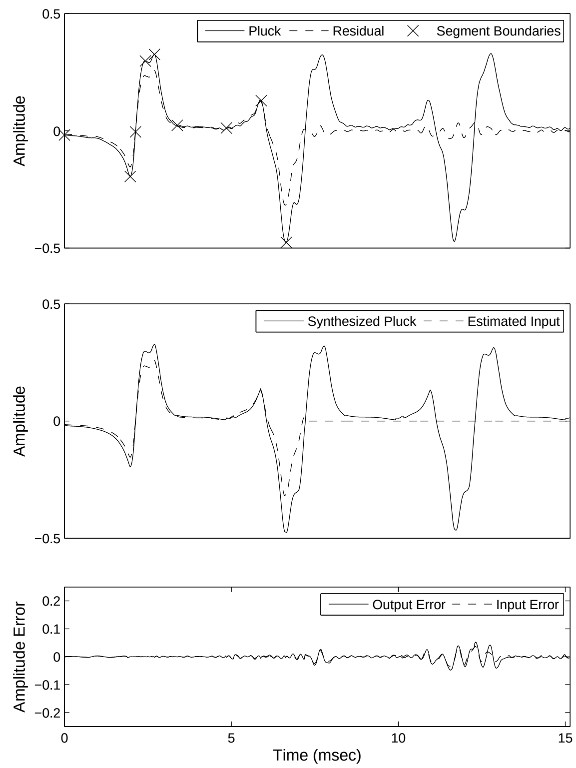

Fig. 5. Analysis and resynthesis of the guitar's 3rd string. Top: Recorded plucked string, residual and excitation boundaries. Middle: Resynthesized pluck and excitation using estimated source-filter parameters. Bottom: Modeling error.

increases with the string's number. The physical explanation for this result is related to the gauge of the string being plucked, which is inversely proportional to the string's number. For example, the 6th string is the thickest on the guitar and exhibits greater pitch fluctuation due to tension modulation when excited. Thus, trying to align a signal exhibiting constant pitch fluctuation with a model based on a fixed pitch will result in sub-optimal source-filter parameter estimates

[1] Sound files are available: http://music.ece.drexel.edu/research/ guitarExpression

Mean-Squared Modeling Error ($\times 10^{-4}$)					
	Pick		Thumb		Average
String	Medium	Soft	Medium	Soft	Error
1	17.0	15.0	20.0	17.0	17.2
2	12.0	6.0	36.0	28.0	20.5
3	14.0	10.0	70.0	10.0	26.0
4	21.0	9.4	29.0	15.0	18.6
5	19.0	11.0	61.0	17.0	27.0
6	40.0	23.0	88.0	17.0	42.0
Average Error	20.5	12.4	50.6	17.3	—

Table 1. Mean-squared error computed between the original and resynthesized plucked-tones.

using the proposed approach. Tension modulation is imparted by the guitarist as well since the strength of their pluck affects the string's maximum displacement. Table 1 demonstrates this, since the model error is greater for the harder plucks.

7. CONCLUSIONS AND FUTURE WORK

We have presented an approach for jointly estimating source and filter parameters of plucked-guitar sounds based on a physically-inspired model. Our approach is motivated by the need to account for variation by both the guitar and the guitarist during the analysis in order to capture expressive intentions at the signal level. Preliminary results indicate that the system is capable of faithfully replicating the perceptual qualities characteristic of various plucking styles by the guitarist when the signals analyzed are well-behaved, that is, they do not exhibit excessive pitch fluctuation.

Future work will entail further refinement of the joint estimation scheme to improve modeling performance on the thicker-gauge strings since the resulting filter parameters were poor fits in these cases. Additionally, we wish to collect pluck samples from many guitarists using the entire fretboard so that we can build quantitative models of expression for many playing styles at all playing positions. Finally, we will validate our analysis-synthesis scheme via formal listening tests to determine if re-synthesized plucking styles have discernible characteristics due to a guitarist's excitation and strength.

8. REFERENCES

[1] K. Karplus and A. Strong, "Digital synthesis of plucked-string and drum timbres," *Computer Music Journal*, vol. 7, no. 2, pp. 43–55, Jun. 1983.

[2] J. O. Smith, "Physical modeling using digital waveguides," *Computer Music Journal*, vol. 16, no. 4, pp. 74–91, 1992.

[3] M. Karjalainen, T. Maki-Patola, A. Kanerva, A. Huovilainen, and P. Janis, "Virtual air guitar," in *117th Audio Engineering Society Convention*. AES, Oct. 2004.

[4] M. Karjalainen, V. Valimaki, and T. Tolonen, "Plucked-string models: From the Karplus-Strong Algorithm to digital waveguides and beyond," *Computer Music Journal*, vol. 22, no. 3, pp. 17–32, Oct. 1998.

[5] V. Valimaki and T. Laakso, "Principles of fractional delay filters," in *Proc. of the 2000 IEEE International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, Jun. 2000, IEEE.

[6] V. Valimaki, J. Huopaniemi, M. Karjalainen, and Z. Janosy, "Physical modeling of plucked string instruments with application to real-time sound synthesis," *Journal of the Audio Engineering Society*, vol. 44, no. 5, pp. 331–353, May 1996.

[7] C. Erkut, V. Valimaki, M. Karjalainen, and M. Laurson, "Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar," in *108th AES Int. Convention 2000*, Paris, France, Feb. 2000, AES, pp. 19–22.

[8] V. Valimaki and T. Tolonen, "Development and calibration of a guitar synthesizer," *Journal of the Audio Engineering Society*, vol. 46, no. 9, pp. 766–778, Sept. 1998.

[9] M. Laurson, C. Erkut, V. Valimaki, and M. Kuushankare, "Methods for modeling realistic playing in acoustic guitar synthesis," *Computer Music Journal*, vol. 25, no. 3, pp. 38–49, Oct. 2001.

[10] M. Karjalainen, V. Valimaki, and Z. Janosy, "Towards high-quality sound synthesis of the guitar and string instruments," in *International Computer Music Conference*. ICMC, Sept. 1993.

[11] G. Cuzzucoli and V. Lombardo, "A physical model of the classical guitar, including the player's touch," *Computer Music Journal*, vol. 23, no. 2, pp. 52–69, Jun. 1999.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, The Edinburgh Building, Cambridge, CB2 8RU, UK, 2004.

[13] L. Rabiner, "On the use of autocorrelation analysis for pitch detection," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 25, no. 1, pp. 24 – 33, feb. 1977.