Techniques for Modeling Expression in Plucked-Guitar Tones

Raymond Migneco and Youngmoo E. Kim
Music and Entertainment Technology Lab (MET-lab)
Electrical and Computer Engineering, Drexel University
{rmigneco, ykim}@drexel.edu

***Abstract*—Source-filter models are a well-established technique for the analysis and synthesis of many acoustic signals, including musical instruments. When applied to the task of modeling plucked-string instruments, these models provide a clear analog to the physical phenomena incurred with exciting the string; that is, an impulsive-like force from the performer excites the resonant behavior of the string. In the case of the guitar, many techniques are available for estimating and calibrating the resonant filter properties of the string, but little research has been invested in the analysis of the driving source signals, which are responsible for reproducing unique timbres associated with the performer's articulation.**

In this work, we present techniques for the analysis and synthesis of plucked guitar signals with emphasis on capturing the expressive attributes of performance. This includes novel approaches for modeling the resonant string behavior and excitation signals of a source-filter model for plucked-guitar tones.

I. Introduction

In recent years, advances in computing have rendered mobile devices and gesture recognition systems cogent platforms for music performance and creation. Touch- and/or gesture-based technologies (e.g. iPad, Kinect) enable entirely new ways of interacting with music. Despite these advances, music creation tools based on these technologies are often use sample-based synthesizers, which limits the expressive capabilities of these systems. Computational models (e.g. physical, source-filter) are capable of simulating the physical characteristics of instruments for expressive control and arbitrary synthesis, especially for the guitar. However, it is unclear how exactly to quantify the nuances of particular playing styles using these models. For music information retrieval applications, such as content-based analysis of performance, computational models of expression are desirable.

This paper presents techniques for the analysis and synthesis of plucked-guitar tones with attention to the expressive attributes of recorded performance. The analysis/synthesis techniques are based on a physically inspired source-filter model, which is discussed in Section II and techniques for calibrating this model are overviewed in Section III. In Section IV, we present a data-driven approach for modeling guitar excitation signals using components analysis. Finally, we present our conclusions in Section V. This work shows that the component analysis techniques applied to guitar recordings can be used to build computational models describing the expressive attributes of the recorded performance.

II. Physically Inspired Modeling

Modeling and synthesis of plucked-guitar tones is often based on digital waveguide (DWG) modeling principles, which digitally implement the d'Alembert solution for traveling waves on a lossy string [1]. The DWG simulates the left- and right-traveling waves occurring after the string is displaced by spatially sampling their time-varying amplitudes along the string's length. It was shown that the DWG model could be reduced to a source-filter interaction as shown in Figure 1 [2].

The lower block, $S(z)$, of Figure 1 is often referred to as the single delay-loop (SDL) and its purpose is to specify the pitch f_0 and model the resonant behavior of the string. The bulk delay filter z^{-D} determines f_0 by the ratio f_s/D where f_s is the audio sampling frequency. The fractional delay filter $H_F(z)$ is used to provide the non-integer delay when required. This creates a resonant filter structure for the fundamental frequency and its harmonics. Since real strings decay with frequency-dependent characteristics, $H_l(z)$ is used to model the decay rates of the harmonics.

The upper block of Figure 1, $C(z)$, is a feedforward comb-filter that incorporates the effect of the performer's plucking point position along the string. Since the SDL lacks the bi-directional characteristics of the DWG, $C(z)$ is required to simulate the boundary reflection when a traveling wave encounters a rigid termination (such as the guitar's nut or bridge). The delay in $C(z)$ is determined by the product βD where β is a fraction in the range $(0, 1)$ corresponding to the relative plucking point location on the string.

An advantage of using the SDL is that it permits model-based analysis of recorded data. This is key in extracting the expressive attributes of performance without having knowledge of the exact physical conditions of the performance.

III. Model Parameter Estimation

Existing techniques for estimating the string model typically involve frequency-domain techniques to determine how the string decays over time. The first step in this process involves determining the fundamental frequency of the string either by autocorrelation function or spectral peak-picking techniques. The short-time Fourier Transform (STFT) is then computed to track how the harmonically-related partials change over time. Using this information, the decay rates for each partial are computed and filter design techniques are used to find

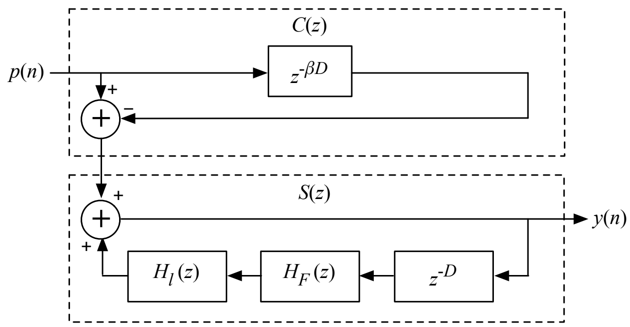


Fig. 1. Source-filter model for plucked-guitar synthesis. $C(z)$ simulates the effect of the player's plucking position along the string. $S(z)$ models the string's resonant behavior and decay characteristics.

a filter with a magnitude response that best matches these specifications [2], [3].

A. Joint Estimation

Rather than employ frequency-based methods to estimate the loop filter of the SDL model, we proposed a novel, time-domain approach for jointly estimating the source and filter parameters from a recorded performance. The SDL output is described as a convolution between the excitation signal and the filter S(z)

$$Y(z) = P_b(z)S(z). \tag{1}$$

$P_b(z)$ is the excitation signal taken at the output of the comb filter $C(z)$ in Figure 1 and contains a bias from the performer's plucking point position (removing this bias is discussed in Section IV).

The parameters for the excitation and filter models can be estimated by minimizing an error term

$$e(n) = p_b(n) - \hat{p}_b(n), \tag{2}$$

where $\hat{p}_b(n)$ is the excitation obtained by filtering the recorded tone with an inverse string model such that $\hat{P}_b(z) = Y(z)S^{-1}(z)$. $p_b(n)$ is a parametric model of the excitation we wish to estimate. In [4], we show that piecewise polynomial functions are suitable for modeling the excitation, which consists of an incident and reflected pulse detected the guitar's pickup.

By using an all-pole filter approximation of $S(z)$, each output sample in $y(n)$ is computed by summing the current excitation sample $p_b(n)$ and a linear combination of previous output samples $y(n-D), y(n-(D+1)), \ldots, y(n-(D+N))$. This allows us to rewrite the minimization in Equation 2 as

$$\mathbf{e} = \mathbf{H}\mathbf{x} - \mathbf{y}, \tag{3}$$

where $\mathbf{H}$ contains past output samples of $y(n)$ and the time indices for each segment of the excitation signal while $\mathbf{x}$ contains the unknown source and filter coefficients. By taking the L_2 norm of Equation 3, a convex optimization problem is formed and $\mathbf{x}$ can be solved using quadratic programming techniques. Additional implementation details and evaluation are presented in [4].

IV. Excitation Modeling

During performance, guitarists convey expression by varying their string articulation in a number of ways. This includes changing the plectrum for displacing the string and the relative dynamics, or strength, used to excite the string. The excitation signals corresponding to these articulations can be recovered from recorded performance by inverse filtering with the SDL model, but it is unclear how these signals should be parameterized to quantify their expressive attributes or for use in expressive synthesis systems. This section overviews a data-driven approach for solving these problems using component analysis techniques.

A. Existing Methods for Excitation Extraction and Modeling

There are several approaches used in the literature for determining the excitation signal for the model shown in Figure 1. One method includes applying non-linear processing to spectrally flatten the recorded tone and use the resulting signal as the source while preserving the signal's phase information [5], [6]. Another technique involves inverse filtering a recorded guitar tone with a properly calibrated string-model [7], [8]. When inverse filtering is used, the string model cancels out the tone's harmonic components related to the fundamental frequency leaving behind a residual that contains the excitation in the first few milliseconds. In [9], these residuals are processed with "pluck-shaping" filters to simulate the performer's articulation dynamics and comb filters to model the reflection.

B. Data Collection and Pre-processing

The analysis of expressive performance for this study consists of particular articulations where the performer varied the dynamics of the articulation and the plucking device. At each fret position, the guitarist performed a specific articulation several times for consistency using either a pick or his finger to excite the string. The neighboring strings are muted so that only the excited string is recorded. Articulations are identified by their dynamic level, which consisted of *piano* (soft), *mezzo-forte* (medium-loud) and *forte* (loud). Approximately 1000 recordings were produced using the first five fretting positions from each of the guitar's 6 strings. A bridge-mounted piezo-electric pickup is used to record the plucked signals because it has a wider frequency response than magnetic pickups and allows the guitarist's plucking point location to be identified using time-domain techniques [10].

The excitation signals are obtained by calibrating an SDL model for each recorded performance and inverse filtering with the associated model. Using the plucking point estimation technique proposed by [10], each excitation signal is "equalized" to remove the comb filter effect which yields a signal closer to a pure impulse.

C. Principal Components Analysis

In previous work, we demonstrated the application of principal components analysis (PCA) to a corpus of excitation signals in order to derive a codebook of basis vectors that can be used to synthesize a multitude of excitation signals

[11]. Here, we briefly overview the application of PCA to the data and discuss how it is used to derive a feature-based representation of the signals in the corpus.

By aligning the excitation signals from our data corpus so that the primary pulse peaks overlap (see Figure 2), we form a data matrix

$$\mathbf{P} = \begin{bmatrix} | & | & & | \\ \mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_N \\ | & | & & | \end{bmatrix}^T \tag{4}$$

where each $\mathbf{p_1}$ is a M-length column vector representing an excitation pulse. The principal components of $\mathbf{P}$ are a set of basis vectors and scores (weights) that can reconstruct the data:

$$\mathbf{P} - \mathbf{u} = \mathbf{W}\mathbf{V}^T. \tag{5}$$

In Equation 5, $\mathbf{u}$ is the mean of $\mathbf{P}$, $\mathbf{V}$ contains the basis vectors of $\mathbf{P}$ along its columns and $\mathbf{W}$ contains the scores (or weightings) to reconstruct each excitation pulse. Several techniques can be used to compute the principal components of $\mathbf{P}$, including the well-known covariance method [12], [13].

Figure 2c plots the first few principal components along with the mean of our data set. The mean vector captures the general impulsive shape of the data, while the components shown serve to widen or narrow the pulse depending on the sign of the associated score value. This relates to the physicality of the string's shape during its initial displacement and finger articulations tend to produce an excitation pulse with greater width than articulations made with a pick. Additional principal components not shown in Figure 2c contribute the noise-like characteristics inherent to the string articulation.

We obtain a feature representation of the excitation signals using the principal components extracted from the data set. By projecting the mean-centered data onto the basis vectors, the principal component scores may be computed as

$$\mathbf{W} = (\mathbf{P} - \mathbf{u})\mathbf{V}. \tag{6}$$

Equation 6 defines an orthogonal linear transformation of the data into a new coordinate system defined by the basis vectors. The scores indicate how much each basis function is weighted when reconstructing the signal. Figure 3a displays the projection of the data onto the first two principal components, which explain the most variance in the data set. We observe that the first principal axis relates to the articulation type (i.e. finger and pick) and strength (e.g. *forte*, *piano*).

D. Nonlinear PCA

Figure 3a shows that there is a nonlinear arrangement of the data when projected onto the principal components. To better relate the expressive attributes of our data set to the component space, we apply nonlinear PCA via autoassociative neural networks (ANN) to the data set of excitation signals.

ANN provides a multi-layer approach for mapping high dimensional data into a lower dimensional space. Features at the input layer of the ANN are transformed by sigmoidal functions in a mapping layer into a lower dimensionality defined by the bottleneck layer [14]. Unlike other nonlinear dimensionality reduction techniques, this process is reversible so the original feature space can be achieved from the bottleneck layer via a de-mapping process. This aspect of ANN's is particularly attractive for our application since we seek a reduced dimensionality space that can quantify the expressive attributes of the data set and be used as a controller for expressive synthesis.

Using the linear PCA scores obtained with Equation 6 as network inputs, we trained an ANN using the Nonlinear PCA MATLAB Toolbox [15]. Empirically, we found that using 25 scores at the input layer of the network was sufficient in terms of adequately describing the data set. As discussed in [11], 25 basis functions explain $> 95\%$ of the variance in the data set and leads to good re-synthesis. Two dimensions were chosen at the bottleneck layer for multiple degrees of freedom, which

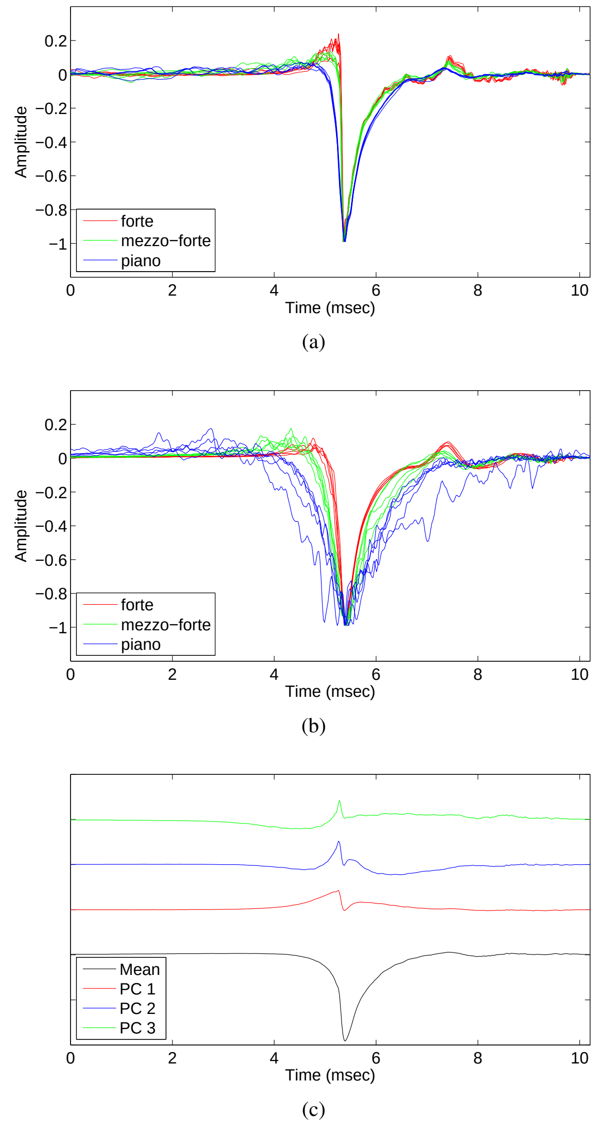
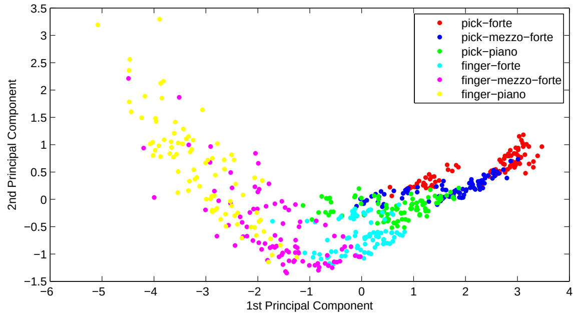
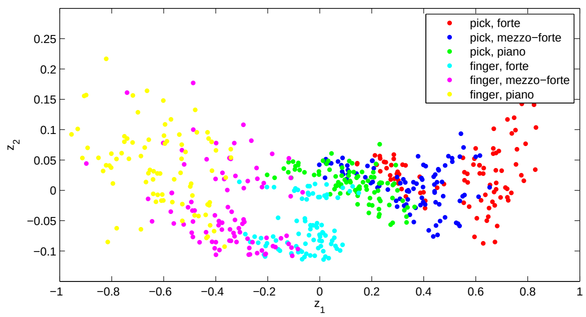


Fig. 2. Example excitation pulses related to articulations produced using (a) a pick and (b) a finger. Principal components extracted from the data are shown in (c) and are offset to highlight their relationship to the pulses in (a) and (b).



(a)



(b)

Fig. 3. (a) Guitar data projected along first two linear component vectors. (b) Projection of data into reduced dimensional space defined by the autoassociative neural network.

could be used to synthesize excitation pulses in an expressive interface.

Figure 3b shows the projection of the data into the reduced dimensionality coordinate space defined by the bottleneck layer of the ANN. Unlike the linear projection shown in Figure 3 (b), the data in the reduced space is clearly distributed around two linear axes. The z_1 axis pertains to articulations produced using either a finger or pick where points sampled in the space $z_1 < 0$ describe finger articulations and points sampled for $z_1 > 0$ pertain to pick articulations. The finger articulations feature a wider excitation pulse in contrast to the pick, where the pulse is generally more narrow and impulsive. In both cases, moving from left to right increases the relative dynamics. The component defined by the z_2 axis relates to the contact time of the articulation. As z_2 is increased, the excitation pulse grows wider for both articulation types. Our informal listening tests confirm that the dimensions of this space correlate with the perceptual qualities of our data corpus when the excitation signals sampled from the space are used to synthesize guitar tones.

V. Conclusion

We have presented techniques for the analysis and synthesis of plucked guitar tones with a focus on quantifying the expressive attributes of recorded performance. By calibrating a source-filter model, we show the excitation signals corresponding to particular string articulations can be derived directly from recordings. Using linear principal components analysis, these signals are characterized by linear basis vectors which relate to the expressive parameters of the data set (i.e. dynamics, plucking device). Finally, nonlinear components analysis was used to achieve a low dimensional representation that compactly describes the attributes of our data set. Future directions for this research include acquisition of additional performance data from a variety of guitarists to expand our computational model for guitar articulations.

Acknowledgment

This research was supported by NSF Award IIS-0644151.

References

[1] J. O. Smith, "Physical modeling using digital waveguides," *Computer Music Journal*, vol. 16, no. 4, pp. 74–91, 1992.
[2] M. Karjalainen, V. Välimäki, 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.
[3] V. Välimäki 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.
[4] R. M. Migneco and Y. E. Kim, "Modeling plucked guitar tones via joint source filter estimation," in *Proc. of the 14th IEEE Digital Signal Processeing Workshop and 6th IEEE Signal Processing Education Workshop (DSP/SPE)*. Sedona, AZ: IEEE, Jan. 2011.
[5] N. Laurenti, G. De Poli, and D. Montagner, "A nonlinear method for stochastic spectrum estimation in the modeling of musical sounds," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 531 –541, Feb. 2007.
[6] N. Lee, Z. Duan, and J. O. Smith, "Excitation signal extraction for guitar tones," in *Proc. of the International Computer Music Conference*. ICMC, 2007.
[7] M. Karjalainen, V. Välimäki, and Z. Janosy, "Towards high-quality sound synthesis of the guitar and string instruments," in *Proc. of the International Computer Music Conference*. ICMC, Sept. 1993.
[8] J. Laroche and J.-L. Meillier, "Multichannel excitation/filter modeling of percussive sounds with application to the piano," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 329 –344, Apr. 1994.
[9] M. Laurson, C. Erkut, V. Välimäki, 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] H. Penttinen and V. Välimäki, "Time-domain approach to estimating the plucking point of guitar tones obtained with an under-saddle pickup," *Applied Acoustics*, vol. 65, pp. 1207–1220, Dec. 2004.
[11] R. M. Migneco and Y. E. Kim, "Excitation modeling and synthesis for plucked guitar tones," in *Proc. of the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, NY: WASPAA, Oct. 2011.
[12] C. Bishop, *Pattern Recognition and Machine Learning*, ser. Information science and statistics. Springer, 2006. [Online]. Available: http://books.google.com/books?id=kTNoQgAACAAJ
[13] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, ser. Pattern Classification and Scene Analysis: Pattern Classification. Wiley, 2001. [Online]. Available: http://books.google.com/books?id=YoxQAAAAMAAJ
[14] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE Journal*, vol. 37, no. 2, pp. 233–243, 1991. [Online]. Available: http://dx.doi.org/10.1002/aic.690370209
[15] M. Scholz, "Nonlinear PCA toolbox for MATLAB," 2011. [Online]. Available: http://www.nlpca.de/matlab.html