EXCITATION MODELING AND SYNTHESIS FOR PLUCKED GUITAR TONES

Raymond V. Migneco and Youngmoo E. Kim

Music and Entertainment Technology Laboratory (MET-lab)
Department of Electrical and Computer Engineering
Drexel University Philadelphia, PA 19104, USA
{rmigneco, ykim}@drexel.edu

ABSTRACT

The analysis and synthesis of plucked-guitar tones via source-filter approximations is a popular and established method for modeling the resonant behavior of the string as well as the driving excitation signal. By varying the source signal, a nearly unlimited number of unique tones can be produced using a given filter model. However, it as unclear as to how exactly the model excitation signals should be parameterized in order to capture the nuances of a guitarist's articulation from a recorded performance. In this paper, we apply principal components analysis to a corpus of excitation signals derived from plucked-guitar recordings in order to design a codebook that captures the unique characteristics of certain string articulations. The development of an excitation codebook has several applications, including expressive synthesis of guitar tones for virtual music interfaces and insight into the expressive intentions of a performer through audio analysis.

***Index Terms*—** Source-filter models, musical instrument synthesis, PCA

1. INTRODUCTION

Source-filter models are a well-established technique for the analysis and synthesis of many acoustic signals, including musical instruments. When applied to 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 [1, 2, 3], but little research has been invested in the analysis of the source signals, which are responsible for reproducing the unique timbres associated with the performer's articulation. The latter problem is complex because there are nearly an infinite number ways to pluck a string, each of which will yield a unique source signal even when the tones have a similar timbre.

In this paper, we apply principal components analysis (PCA) to a corpus of excitation signals derived from recordings of plucked-guitar tones in order to derive parameters useful for modeling the unique characteristics of guitar articulations. As we will discuss, PCA is employed for this task in order to exploit the common features of the excitation signals while modeling the finer details using the appropriate principal components. Our approach can be viewed as designing a codebook where the entries are the components that describe the unique characteristics of the excitation signals. This research has several applications, including modeling guitar performance directly from recordings in order to capture expressive and perceptual characteristics of a performer's playing style. Additionally, the codebook entries obtained in this paper can be applied to musical interfaces for control and synthesis of expressive guitar tones.

2. BACKGROUND

Synthesis of plucked-guitar tones is often based on digital waveguide (DWG) modeling principles, which were introduced by Smith to simulate the d'Alembert solution for traveling waves on a lossy string [4]. The DWG model for a guitar utilizes two spatially sampled delay lines to model the time-varying amplitudes and positions of the left- and right-traveling wave shapes that result from releasing an initially displaced string. Later, Karjalainen et al. showed that the DWG model could be reduced to the so-called single-delay loop (SDL) model, which is shown in Figure 1 [5]. This model consolidates the components of the DWG model into a single delay line z^{-D_I} in cascade with a loop filter $H_l(z)$ and a fractional delay filter $H_F(z)$. The loop and fractional delay filters are calibrated such that the total delay D in the SDL satisfies $D = \frac{f_s}{f_0}$ where f_s and f_o indicate the sampling frequency and target pitch, respectively. $H_l(z)$ is designed by determining a filter with a magnitude response matching the decay rates for the harmonically-related partials in the tone [1, 2, 3].

While the SDL is essentially a source-filter approximation of the physical system for a plucked-string, there are several benefits associated with modeling tones in this manner. For example, modifying the source signal permits arbitrary synthesis of unique tones even for the same filter model. Also, for analysis tasks it is desirable to model the perceptual characteristics of tones from a recorded performance by recovering the source signal using linear filtering operations, which is possible with a source-filter model.

There are several approaches used in the literature for determining the excitation signal for the source-filter model of a plucked-guitar. A possible source signal includes filtered white noise, which simulates the transient noise-like characteristics of a plucked-string. Other methods utilize non-linear processing to spectrally flatten the recorded tone and use the resulting signal as the source, since it preserves the signal's phase information [6, 7]. A well-known technique involves inverse filtering a recorded guitar tone with a properly calibrated string-model [1, 2]. When inverse filtering is used, the string model cancels out the tone's harmonic components leaving behind a residual that contains the excitation in the first few milliseconds. In [8], these residuals are processed with "pluck-shaping" filters to simulate the performer's articulation dynamics. For improved reproduction of acoustic guitar tones, this approach is extended by decomposing the tone into its deterministic and

978-1-4577-0693-6/11/$26.00 ©2011 IEEE

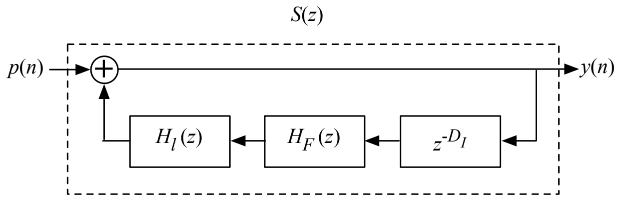

Figure 1: Single delay-loop (SDL) model for plucked-string synthesis.

stochastic components, separately inverse filtering each signal and adding the residuals to equalize the spectra of the residual [9].

3. EXCITATION MODELING

It is well known by guitarists that exactly reproducing a particular articulation on a guitar string is extremely difficult, if not impossible. However, we hypothesize that excitation signals obtained from analyzing similarly plucked-guitar tones exhibit common characteristics that can be parametrically represented. Techniques used by guitarists to vary their articulation include altering the dynamics (i.e. relative "hardness" or "softness") of their picking technique and also changing the mechanism for exciting the string (i.e. using a pick, nail or finger). These techniques have a direct impact on the initial shape of the string, and yield perceptually unique timbres especially during the "attack" instant of the tone.

3.1. Recovering the Excitation

To demonstrate the signal-level differences associated with different articulations, we recover the excitation signals for different plucked-guitar tones produced by using either a pick or the fleshy part of a finger to produce the articulation. These excitations were recovered by inverse filtering the recorded tone with a string filter so that

$$P(z) = Y(z)/S(z) \tag{1}$$

where $P(z)$, $Y(z)$, and $S(z)$ represent the excitation signal, recorded tone and string model, respectively. We employ the SDL model described in Section 2 for our string filter. The recovered excitations were normalized and aligned by using the lag computed from the cross-correlation between each $p(n)$ and the same reference signal to ensure the significant features overlapped. After alignment, the first 15 milliseconds of each residual is retained to avoid including amplitude and phase errors from the string filter. As our corpus consists of electric guitar recordings, truncating the residual does not omit the body resonance effects observed in acoustic guitars [9].

Figure 2 shows that the recovered signals share the same contours, which are related to the d'Alembert solution for a plucked-string since observing the sum of the left- and right-traveling disturbances at a particular location along the string yields patterns of constructive or destructive interference for each period of vibration [10]. The deviations between these signals are also obvious as well. In particular, residuals corresponding to pick articulations exhibit sharp, impulse-like transitions near areas of interference, where the the finger articulations are generally smoother in these areas. Thus, the residuals reflect the impact of the performer's finger or pick on the string's initial shape.

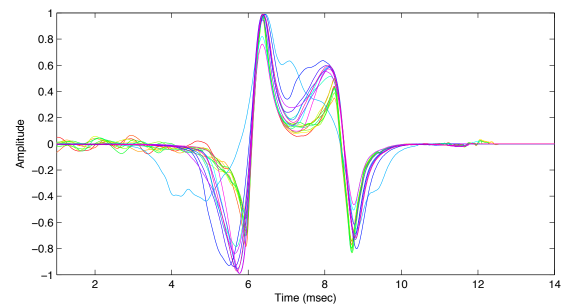

Figure 2: Over-layed excitation signals obtained by inverse filtering recorded tones produced by varying the string articulation.

3.2. PCA Modeling

We investigate modeling the excitation signals obtained from plucked-guitar tones using principal components analysis (PCA) since we wish to exploit the common characteristics of our signals while modeling the details with only the essential principal components required for a particular set of articulations.

We perform PCA on our data by forming a matrix $\mathbf{P} = [\mathbf{p}_1\mathbf{p}_2 \ldots \mathbf{p}_N]$ where each $\mathbf{p}_i$ is a column vector containing a M-dimensional signal from the N signals in our database. The residual signals are recovered from recordings sampled at 44.1kHz and we choose $M = 650$ as the normalized length to yield a duration of $\sim$15 milliseconds for each signal. A particular excitation can be represented by taking the linear combination

$$\mathbf{x}_i = \mathbf{p}_i - \bar{\mathbf{p}} = \sum_{m=1}^{M} w_{i,m}\mathbf{v}_m \tag{2}$$

where $\bar{\mathbf{p}}$ is the empirical mean of our data and $\mathbf{v}$ are the eigenvectors of the covariance matrix computed by taking $E(\mathbf{x}\mathbf{x}^T)$. The PCA weights $\mathbf{w}$ are obtained by projecting $\mathbf{x}$ onto the components in $\mathbf{v}$ [11]. The eigenvectors in $\mathbf{v}$ are arranged corresponding to the decreasing order of the associated eigenvalues so that $\lambda_1 > \lambda_2 > \cdots > \lambda_M$.

PCA provides us with a means of determining how many dimensions are needed to accurately and compactly represent our data. We examine compactness by calculating the explained variance (EV), which is obtained by selecting $M' < M$ and computing $\Sigma_m^{M'}\lambda_m/\Sigma_m^M\lambda_m$. In Figure 3 we see that selecting $M = 20$ accounts for 99% of the EV, which suggests that the dimensionality of the excitation signals is relatively low ($\ll 650$).

Plotting the mean vector $\bar{\mathbf{p}}$ along with the first 3 principal components $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ provides additional insight on how PCA decomposes our data. In Figure 4 we see that $\bar{\mathbf{p}}$ captures the general contour of the excitations plotted in Figure 2 while the principal components provide the finer details of the articulations by incorporating high-frequency information into the model.

3.3. Codebook Design

Though 99% of our data's variance may be explained by taking $M = 20$ principal components (PCs), we expect that only a subset of these are required to reconstruct the signals with reasonable accuracy. Furthermore, we do not expect that these subsets are necessarily the same for finger or pick articulations, since their corre-

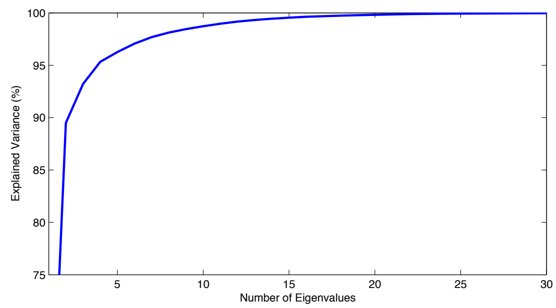

Figure 3: Proportion of explained variance achieved by varying the number of principal components used in the analysis.

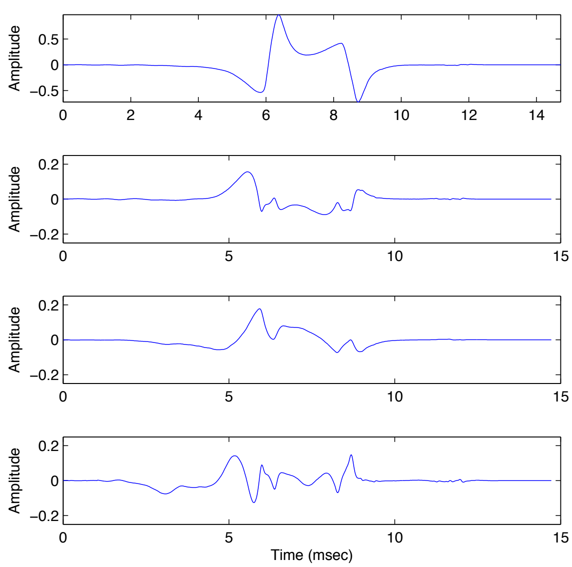

Figure 4: Mean vector $\bar{\mathbf{p}}$ (top) of the extracted excitation signals followed by the first 3 principal components.

sponding excitation signals have distinguishing features as shown in Figure 2. Thus, we seek the essential PCs to comprise codebooks that describe excitation signals corresponding to a particular articulation.

To determine the subsets of PCs comprising our codebook, we sort the first 20 PC weights for each signal in descending order to obtain $\mathbf{w} = [w_1 w_2 \ldots w_{20}]$ where $w_1 > w_2 > \cdots > w_{20}$. For each signal, the 5 largest weights in $\mathbf{w}$ are selected and used to determine the distribution of the top PC weights for all the signals corresponding to a particular articulation. Figure 5 shows the distributions of these top PC weights corresponding to the finger and pick excitations. From Figure 5, it is evident that PC 1 is the top weight for pick articulations and thus an essential component for representing the associated excitation signals. On the other hand, the top weights for finger excitations are approximately uniformly distributed over PCs 2 through 11.

Using the data in Figure 5 as a guide, the codebook entries for each articulation type are selected by choosing L of the top PC components with the largest weights. This yields a subset of weights $\hat{w}$

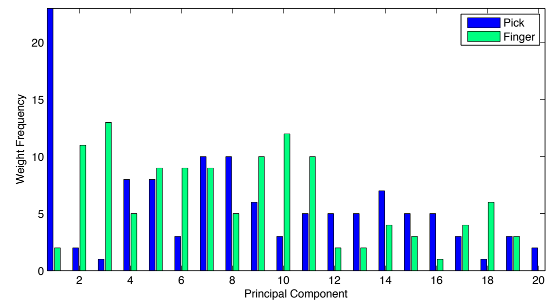

Figure 5: Weight frequencies for the first 20 principal components for finger and pick articulations.

and associated eigenvectors $\hat{\mathbf{v}}$ from the initial PC analysis where $\hat{w}_i \subset w_i$ and $\hat{\mathbf{v}} \subset \mathbf{v}$, respectively. The first 5 codebook entries for pick articulations are shown in Figure 6, where it is evident that the PCs incorporate increasing high frequency detail into the source model.

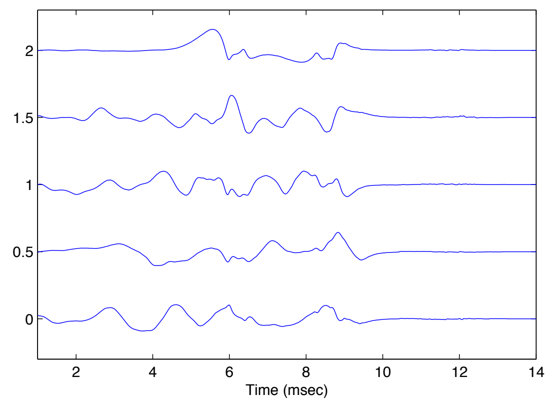

Figure 6: 5 codebook entries for excitation signals corresponding to guitar tones produced with a pick. Entries are offset for clarity.

3.4. Excitation Synthesis

By defining the subsets of PC components as entries for the codebooks representing finger and pick articulations, the excitation signals are reconstructed by taking

$$\mathbf{x}_i = \bar{\mathbf{p}} + \sum_{l=1}^{L} \hat{w}_{i,l} \hat{\mathbf{v}}_l. \tag{3}$$

In (3), $\hat{w}_{i,l}$ and $\hat{\mathbf{v}}_l$ are the subsets consisting of L weights and eigenvectors from the original PC data describing each articulation.

To evaluate the quality of the reconstruction, we compute the signal-to-noise ratio (SNR) between the original excitations and the reconstructions generated by varying the number of codebook entries used in (3). The average SNR values for pick and finger articulations are reported in Table 1, which reveals some interesting trends. Notably, the pick articulations are well-represented by a few codebook entries with moderate SNR improvement as more entries are added. On the other hand, the SNR improves significantly as

	SNR (dB) for Number of Codebook Entries									
Pluck Type	1	2	3	4	5	6	7	8	9	10
Pick	17.78	18.05	18.18	18.35	18.81	18.90	19.01	19.10	19.24	19.40
Finger	10.55	10.56	13.30	13.35	13.44	13.60	13.87	13.93	13.94	14.62

Table 1: Average signal-to-noise ratio computed between recovered and synthetic excitations. The columns indicate the SNR corresponding to the number of codebook entries used for synthesis.

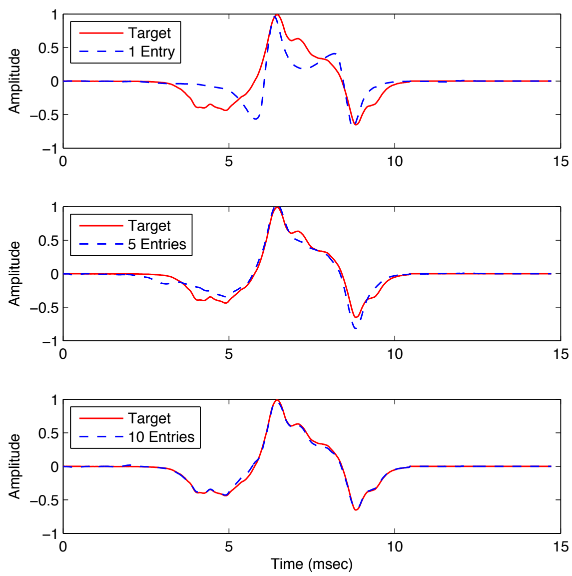

Figure 7: Reconstructed excitation signal produced by varying the number of codebook entries.

more entries are used for finger articulations, which suggests using additional codebook entries to better represent these signals. This result is not surprising, since we observed more irregularities between different finger articulations, which leads us to believe that it is more difficult to control exactly how the finger separates from the string.

Figure 7 demonstrates the quality of reconstruction achieved by varying the number of codebook entries for a finger excitation signal. Clearly, the complex contours of the signal are not well modeled using one entry, but a reasonably good approximation is achieved with ten entries.

4. DISCUSSION AND FUTURE WORK

This paper presents a novel approach towards modeling the excitation signals that drive a source-filter model for plucked-guitar tones using PCA. PCA is capable of modeling the common features of the excitation signals, which are related to the physical behavior of plucked strings, as well as the fine details that distinguishes the type of articulation used by the performer. Additionally, a framework for extracting a codebook that describes a particular articulation is presented, which permits reasonably good approximations of the original excitation signals using only a few principal components.

Future work entails further analysis of plucked-guitar tones to develop excitation codebooks that describe additional dimensions of expressiveness used by the performer, such as dynamics and additional picking devices. Furthermore, a perceptual analysis on the synthetic guitar tones created with the codebooks should be investigated to determine the model's validity.

5. ACKNOWLEDGMENT

This work is supported by NSF Award IIS-0644151.

6. REFERENCES

[1] 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.

[2] J. Laroche and J.-L. Meillier, "Multichannel excitation/filter modeling of percussive sounds with application to the piano," *Speech and Audio Processing, IEEE Transactions on*, vol. 2, no. 2, pp. 329 –344, Apr. 1994.

[3] 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.

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

[5] 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.

[6] N. Laurenti, G. De Poli, and D. Montagner, "A nonlinear method for stochastic spectrum estimation in the modeling of musical sounds," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 531 –541, Feb. 2007.

[7] N. Lee, Z. Duan, and J. O. Smith III, "Excitation signal extraction for guitar tones," in *Proc. of the 2007 International Computer Music Conference*. ICMC, 2007.

[8] 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.

[9] 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.

[10] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. Wiley, 1982.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley, 2001.