A Component-Based Approach for Modeling Plucked-Guitar Excitation Signals

Raymond V. Migneco
Music and Entertainment Technology Laboratory (MET-lab)
Dept. of Electrical and Computer Engineering
Drexel University Philadelphia, PA 19104, USA
rmigneco@drexel.edu

Youngmoo E. Kim
Music and Entertainment Technology Laboratory (MET-lab)
Dept. of Electrical and Computer Engineering
Drexel University Philadelphia, PA 19104, USA
ykim@drexel.edu

ABSTRACT

Platforms for mobile computing and gesture recognition provide enticing interfaces for creative expression on virtual musical instruments. However, sound synthesis on these systems is often limited to sample-based synthesizers, which limits their expressive capabilities. Source-filter models are adept for such interfaces since they provide flexible, algorithmic sound synthesis, especially in the case of the guitar. In this paper, we present a data-driven approach for modeling guitar excitation signals using principal components derived from a corpus of excitation signals. Using these components as features, we apply nonlinear principal components analysis to derive a feature space that describes the expressive attributes characteristic to our corpus. Finally, we propose using the reduced dimensionality space as a control interface for an expressive guitar synthesizer.

Keywords

Source-filter models, musical instrument synthesis, PCA, touch musical interfaces

1. INTRODUCTION

In recent years, advances in computing have rendered mobile devices and gesture recognition systems cogent platforms for music performance and creation. Devices such as the iPad and Kinect enable touch- and/or gesture-based interaction to enable entirely new ways of interacting with music. Despite these advances, the software on these systems still relies heavily on sample-based synthesizers, which limits the expressive control available to the user. Source-filter models are capable of simulating the physical characteristics of plucked-string instruments, including the resonant string behavior. Unlike sample-based synthesizers, these models can generate a wide range of musical timbres in response to different excitation signals. However, it is unclear how exactly the source signals should be modeled to capture the nuances of particular playing styles.

In this paper, we explore the analysis and synthesis of plucked-guitar tones via component analysis of residual signals extracted from recorded performance for the application of expressive guitar synthesis. The rest of this paper is as follows: In Section 2 we briefly overview physically inspired modeling of plucked-guitar tones along with existing methods for modeling excitation signals. Section 3 describes our data set and how the excitation signals are extracted from recorded performance. In Section 4 we obtain a feature representation of our signals using principal components analysis and apply non-linear components analysis to these features for dimensionality reduction in Section 5. Finally, in Section 6 we demonstrate an interface for expressive guitar synthesis using the reduced dimensionality space.

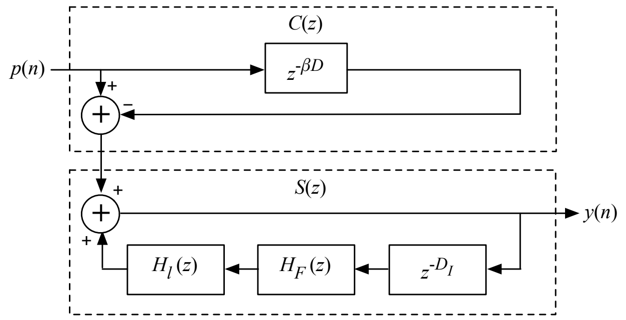


Figure 1: Source-filter model for plucked-guitar synthesis. $C(z)$ simulates the effect of the player's plucking position. $S(z)$ models the string's pitch and decay characteristics.

2. BACKGROUND

Modeling and synthesis of plucked-guitar tones is often based on digital waveguide (DWG) modeling principles, which aim to digitally implement the d'Alembert solution for traveling waves on a lossy string [18]. The DWG simulates the left- and right-traveling waves occuring after the string is displaced by spatially sampling their time-varying amplitudes along the string's length. It was later shown that the DWG model could be reduced to a source-filter interaction as shown in Figure 1 [7]. The lower block, $S(z)$, of Figure 1 is referred to as the single delay-loop (SDL) and consolidates the DWG model into a single delay line z^{D_I} in cascade with a string decay filter $H_l(z)$ and a fractional delay filter $H_F(z)$. These filters are calibrated such that the total delay, D, in the SDL satisfies $D = \frac{f_s}{f_0}$ where f_s and f_0 are the sampling frequency and fundamental frequency of the tone, respectively. The upper block, $C(z)$, is a feed-forward 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, $C(z)$ simulates the boundary conditions when a traveling wave encounters a rigid termination. 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.

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 using the resulting signal as the source while preserving the signal's phase information [10, 12]. Another technique involves inverse filtering a recorded guitar tone with a properly calibrated string-model [6, 9]. 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 [11], these residuals are processed with "pluck-shaping" filters to simulate the performer's articulation dynamics and comb filters to model the reflection.

By employing the waveguide principles for plucked-string synthesis, Karjalainen et al. developed a Virtual Air Guitar interface for expressive performance [5]. The system utilized sensors worn on the performer's hands in order to determine specific playing gestures such as plucking, strumming, vibrato and pitch. However, the signals used to excite the filter model are limited to stored residual signals obtained by inverse filtering recorded guitar performance. More recently, the open source community has employed gesture tracking technology used in the Microsoft Kinect to develop a controller-free air guitar interface [14]. While this system relies on sample-based and not algorithmic sound synthesis, it provides a compelling interface for capturing the performer's expression.

Recently, a variety of virtual guitar applications have been developed for the iPad and integrate some degree of expressive control over the resulting sound. Among these are iPad's implementation of Garageband, which uses accelerometer data in response to the user's tapping strength to trigger an appropriate sample for the synthesizer [2]. Similarly, the OMGuitar enables single note or chorded performance and triggers chord samples based on the how quickly the user "strums" the interface [1]. In both cases, sound synthesis is based on samples of recorded guitars.

3. DATA COLLECTION

Our data corpus consists of recordings produced using an Epiphone Les Paul guitar equipped with a Fishman Powerbridge pickup. This pickup is a modified bridge with piezoelectric sensors installed in the saddles for each string. In contrast to magnetic pickups, the piezo pickup responds to pressure changes caused by string vibration at the bridge. These pickups provide a wide frequency response, which is desirable for modeling the noise-like characteristics of the performer's articulation. Furthermore, these pickups do not include the low-pass characteristics incurred from magnetic pickups and are relatively free of the resonant effects from the guitar body. Finally, recordings obtained through the bridge-mounted piezo pickup can be analyzed to determine the guitarist's plucking position along the string since the output is always measured at the bridge.

The data set of plucked-guitar recordings was produced by varying the articulation to produce different notes using various positions on the fretboard including "open" strings. At each fret position, the guitarist performed a specific articulation several times for consistency using either the pick or his finger to excite the string. The neighboring strings are muted so that only the excited string is recorded by the pickup. Articulations are identified by their dynamic level, which consisted of *piano* (soft), *mezzo-forte* (medium-loud) and *forte* (loud). All six strings were used including the first

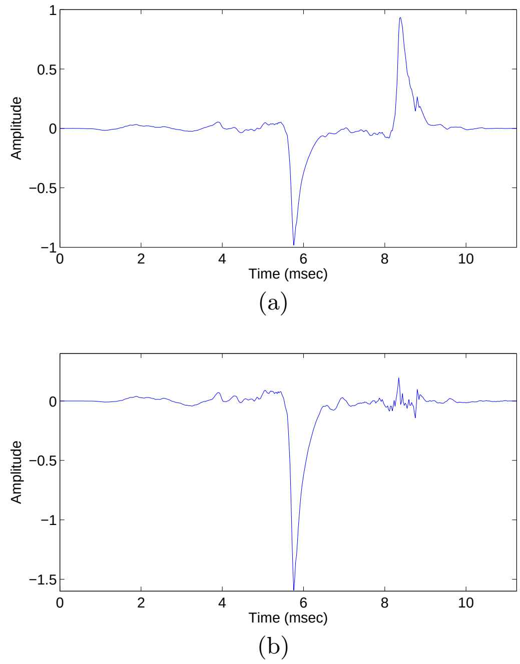


Figure 2: Plucking point compensation for a residual signal obtained from plucking a guitar string 8.4 cm from the bridge (open E, $f_0 = 331$ Hz). (a) Without and (b) with equalization

.

five fretting positions to yield approximately 1000 recordings. The output of the pick-up was fed to a M-Audio Fast Track Pro USB interface, which recorded the audio directly to a Macintosh computer running Audacity. Samples were recorded at 44.1 kHz at a 16-bit depth.

3.1 Residual Extraction

We obtain residual excitation signals from our data by inverse filtering the recorded tone with a properly calibrated SDL model. The techniques proposed in [6, 9, 20] were used to calibrate the SDL parameters. However, the residual obtained by inverse filtering contains a bias from the comb-filter effect resulting from the guitarist's plucking position along the string. In the frequency domain, this residual will contain deep notches at the harmonics related to the plucking position. Since the plucking point position typically varies in real performance and in our data set, it must be compensated for to standardize the analysis. We employ a technique proposed by Penttinen et al. developed to estimate the relative plucking position on guitars equipped with bridge-mounted pickups [15]. The relative plucking position is used to calibrate the comb filter $C(z)$ in Figure 1 to remove the deep spectral notches.

The total inverse filtering operation of the recorded signal is then expressed as

$$P(z) = \frac{Y(z)}{C(z)S(z)}. \tag{1}$$

Figure 2 shows a residual excitation signals before and after the comb filter effect is removed. Besides standardizing the analysis, removing the comb-filter effect allows the relative plucking point position to remain a free parameter for re-synthesis. It should be noted that, in the compensated case, the excitation pulse approaches an ideal impulse. This is related to the piezoelectric sensor responding to acceleration rather than displacement, which is the wave variable most often used in DWG models [18].

4. PCA FEATURES

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 synthesize a multitude of excitation signals [13]. 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.

4.1 Principal Components Analysis

Since the pulse widths are dependent to some degree on the fundamental frequency of the string, we first normalize all the pulses to a common period. The signals are then aligned in the time domain so that the primary peak of the pulses overlap as shown in Figure 3. Using the aligned signals, a data matrix is constructed

$$\mathbf{P} = \begin{bmatrix} | & | & & | \\ \mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_N \\ | & | & & | \end{bmatrix}^T \tag{2}$$

where each $\mathbf{p}$ 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:

$$\mathbf{P} - \mathbf{u} = \mathbf{W}\mathbf{V}^T. \tag{3}$$

In Equation 3, $\mathbf{u}$ 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 [3, 4].

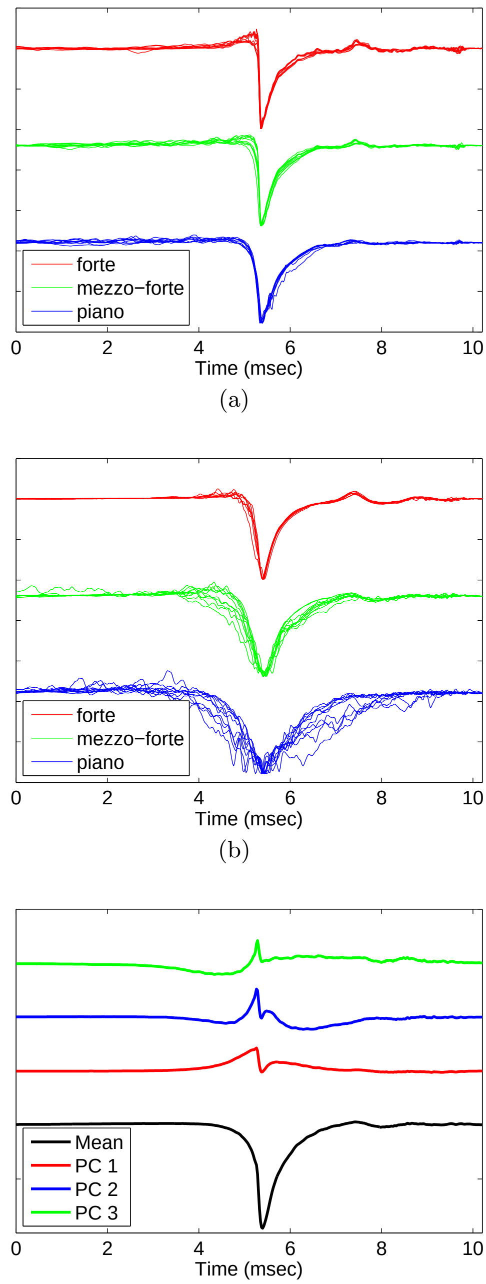
Figure 3(c) plots the first few principal components along with the mean of our data set. The mean vector captures the general impulsive shape of the data, while the components shown serve to widen or narrow the pulse depending on the sign of the associated score value. This relates to the physicality 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 3(c) contribute the noise-like characteristics inherent to the string articulation. The number of basis vectors obtained via PCA is equivalent to the number of variables used to model the data. In this case, 570 vectors comprise $\mathbf{V}$, however, in [13] we show that using a subset of the basis vectors is sufficient for re-generating the pulse with good accuracy.

4.2 Feature Representation

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

$$\mathbf{W} = (\mathbf{P} - \mathbf{u})\mathbf{V}. \tag{4}$$

Equation 4 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 4 displays the projection of the data onto the first two principal components since this pair of components explains the most variance in the data. We observe that the first principal axis relates to the articulation type (i.e. finger and pick) and strength (e.g. *forte*, *piano*). However, due to the nonlinear distribution of the data along these axes, it is unclear how these and additional components exactly relate to the properties of the excitation pulses.



(c)

Figure 3: Example pulses related to articulations produced using (a) pick and (b) finger to excite the string. Principal components extracted from the data are shown in (c) and are offset to highlight their relationship to the pulses in (a) and (b).

5. NONLINEAR PRINCIPAL COMPONENTS ANALYSIS

While the linear PCA technique presented in the previous section provides insight on the underlying basis functions comprising our data set, it is unclear how the high dimensional component space relates to the expressive attributes of our data. As shown in Figure 4, there is an underlying nonlinear distribution of the data along the principal axes. In this section, we apply nonlinear principal components analysis (NLPCA) to the scores extracted from linear PCA to derive a lower dimensional representation of the data.

5.1 Background

There are many techniques available in the literature for nonlinear dimensionality reduction, or *manifold-learning*, for the purposes of discovering the structure of high dimensional, nonlinear data. Such techniques include *locally linear embedding* (LLE) [16] and *Isomap* [19]. While LLE and Isomap are useful for data reduction and visualization tasks, their application does not provide an explicit mapping function to project the reduced dimensionality data back into the high dimensional space.

For the purpose of developing an expressive control interface, re-mapping the data back into the original space is essential since we wish to use our linear basis vectors to reconstruct the excitation pulses. To satisfy this requirement, we employ NLPCA via autoassociative neural networks (ANN) to achieve dimensionality reduction with explicit re-mapping functions.

The standard architecture for an ANN is shown in Figure 5 and consists of 5 layers [8]. The input and mapping layers can be viewed as the "extraction" function since it projects the input layers into a lower dimensional space as specified in the bottleneck layer. The de-mapping and output layers comprise the "generation" function, which projects the data back into its original dimensionality. Using Figure 5 as an example, the ANN can be specified as a 3-4-1-4-3 network to indicate the number of nodes at each layer. The nodes in the mapping and de-mapping functions contain sigmoidal functions and are essential for compressing and decompressing the range of the data to and from the bottle neck layer. Since the desired values at the bottleneck layer are unknown, direct supervised training cannot be used to learn the mapping and de-mapping functions. Rather, the combined network is learned using back propagation algorithms to minimize a squared error criterion such that $E = \frac{1}{2}\|\mathbf{w} - \hat{\mathbf{w}}\|$ [8]. From a practical standpoint, this yields a set of transformation matrices to compress (T_1, T_2) and decompress (T_3, T_4) the dimensionality of the data.

5.2 Application to Guitar Data

To uncover the nonlinear structure of the guitar features extracted in Section 4.2, we employed the NLPCA MATLAB Toolbox to extract our ANN [17]. Empirically, we found that using 25 scores at the input layer was sufficient in terms of adequately describing the data set and expediting the ANN training. As discussed in [13], 25 basis functions explain $> 95\%$ of the variance in the data set and leads to good re-synthesis. At the bottleneck layer of the ANN,

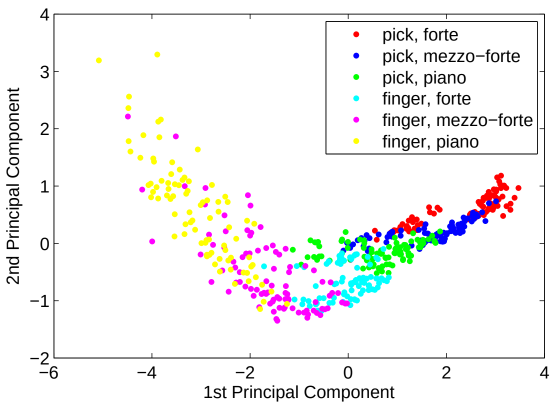


Figure 4: Projection of guitar excitation signals along the first two principal axes.

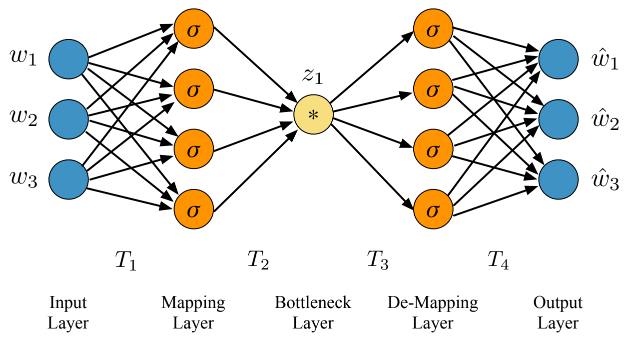


Figure 5: Example autoassociative neural network with 3-4-1-4-3 architecture.

we chose two nodes in order to have multiple degrees of freedom which could be used to synthesize excitation pulses in an expressive interface. These design criteria yielded a 25-6-2-6-25 ANN architecture.

Figure 6 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 4, the data in the reduced space is clearly clustered around the z_1 and z_2 axes. Selected excitation pulses are also shown, which were synthesized by sampling this coordinate space, projecting back into the linear principal component domain using the transformation matrices (T_3, T_4) from the ANN and using the resulting scores to reconstruct the pulse with linear component vectors.

The nonlinear component defined by the z_1 axis describes the articulation type where points sampled in the space $z_1 < 0$ pertain to 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 articulation spaces, moving from left to right increases the relative dynamics. The second nonlinear 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.

6. INTERFACE

We demonstrate the practical application of this research in a touch-based iPad interface shown in Figure 7. This interface acts as a "tabletop" guitar, where the performer uses one hand to provide the articulation and the other to key in the desired pitch(es). The articulation is applied to the large, gradient square in Figure 7, which is a mapping of the reduced dimensionality space shown in Figure 6. Moving up along the vertical axis of the articulation space increases the dynamics of the articulation (*piano* to *forte*) and moving right to left on the horizontal axis increases the contact time. The articulation area is capable of multi-touch input so the performer can use multiple fingers within the articulation area to give each tone a different timbre.

The colored keys on on the left-side of Figure 7 allow the user to produce certain pitches. Adjacent keys on the horizontal axis are tuned a half step apart and their color indicates that they are part of the same "string" so that only the leading key on the string can be played at once. Diagonal keys on adjacent strings are tuned to a Major 3rd interval while the off-diagonal keys represent a Minor 3rd interval. This arrangement allows the performer to easily finger different chord shapes.

The synthesis engine for the tabletop interface must is ca-

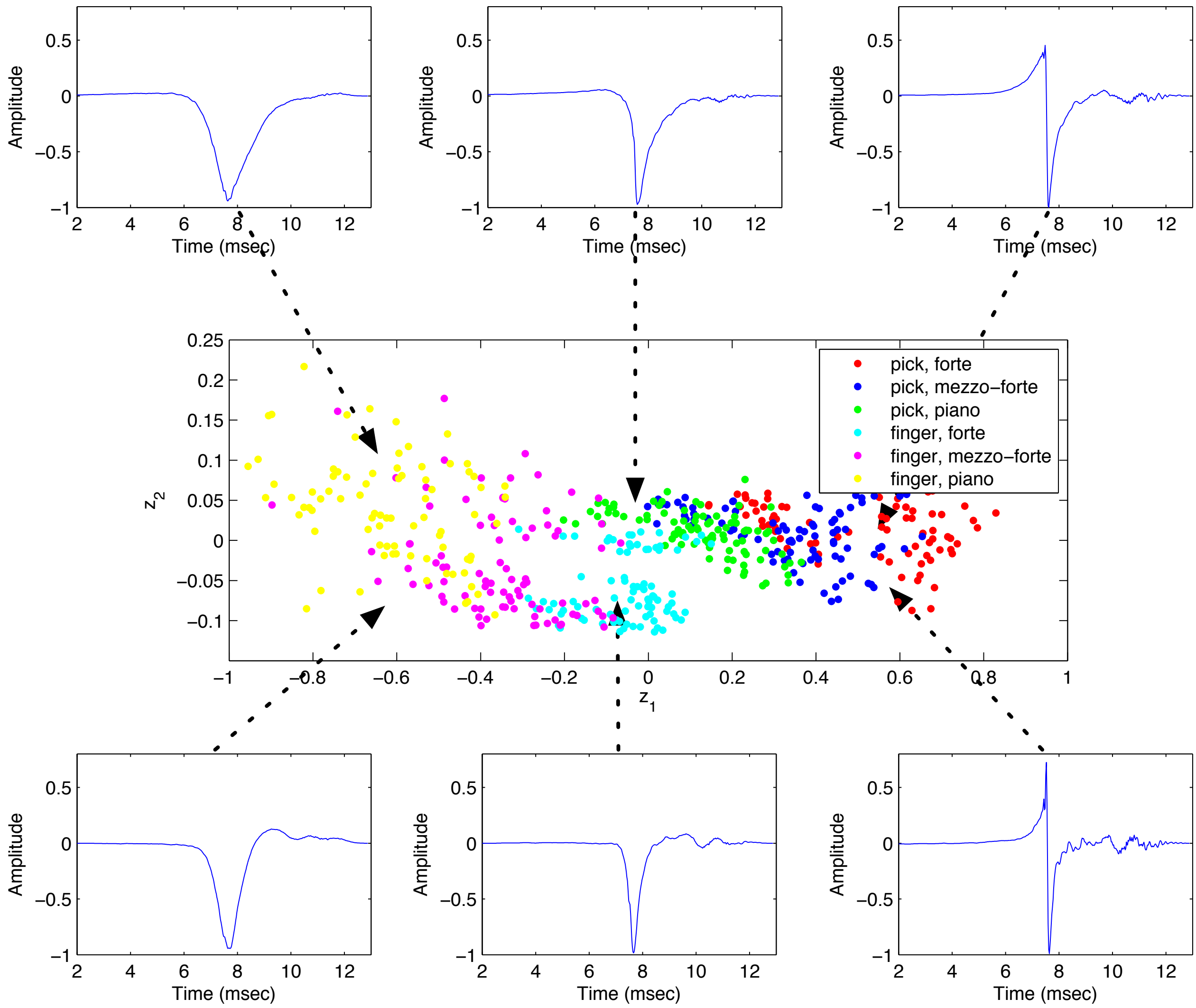


Figure 6: Projection of the guitar data into the reduced dimensionality space defined by the ANN (center). Example excitation pulses resulting from sampling this space are also shown.

pable of computing the excitation signal corresponding to the performer's touch point within the articulation space and filtering the resulting excitation signal for multiple tones in real-time. The filter module used for the string is implemented with the single delay-loop model shown in Figure 1. Though this filter has a large number of delay taps, which is dependent on the pitch, only a few of these taps have non-zero coefficients, which permits an efficient implementation of infinite impulse response filtering. Currently, the relative plucking position along the string is fixed, though this may be a free parameter in future versions of the application. The excitation signal can be updated in real-time during performance, which is made possible by the iPad's support of hardware-accelerated vector libraries. These include the matrix multiplication routines to project the low dimensional user input into the high dimensional component space. Through our own testing, we found that the excitation signal is typically computed in < 1 millisecond, which is more than adequate for real-time performance.

7. CONCLUSIONS

We have presented a novel approach for modeling the excitation signals for plucked-guitar tones using principal components analysis. Our method draws on physically inspired modeling techniques to extract the excitation pulses from recorded performances pertaining to various articulation styles. Using linear principal components analysis, these excitation signals are modeled by a set of linear basis vectors. The associated weights for these basis vectors are then used as features to train an autoassociative neural network, which provides a nonlinear mapping to a reduced dimensionality space. By sampling points in the reduced dimensionality space, we show that a wide range of excitation pulses can be synthesized, which correlate to the expressive attributes of our data corpus, namely articulation type, strength and contact time. We have also demonstrated the practical application of this research by implementing the excitation and plucked-string synthesis into an iPad application, which is capable of real-time guitar synthesis with control over the expressive attributes in our data set.

As demonstrated with the iPad application, this research is extremely applicable to virtual instrument technology. Beyond touch interfaces, it may be possible to leverage ges-

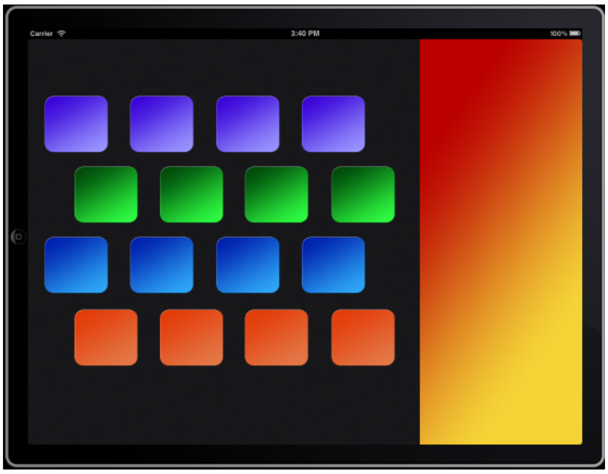


Figure 7: Tabletop guitar interface for the components based excitation synthesis. The articulation is applied in the gradient rectangle, while the colored squares allow the performer to key in specific pitches.

ture recognition, such as the Microsoft Kinect, to trigger particular articulations. By freeing the user from the constraints of a physical device, a unique-gesture based synthesizer could be built for "air-guitar" applications.

Avenues for further research include the acquisition of additional performance data from a variety of guitarists. This data collection and subsequent analysis could lead to computational models describing the stylings of particular performers. These models could be used to "profile" particular players and integrate their stylings into virtual music interfaces. From a physical modeling standpoint, the guitar synthesis model used in our application can be expanded to include magnetic pickups and resonant body effects, which factor into perceived timbres of real acoustic and electric guitars.

8. ACKNOWLEDGMENTS

This research was supported by NSF Award IIS-0644151.

9. REFERENCES

[1] Amidio. Omguitar advanced guitar synth, Jan. 2012.

[2] Apple. Garageband, Jan. 2012.

[3] C. Bishop. *Pattern Recognition and Machine Learning*. Information science and statistics. Springer, 2006.

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

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

[6] 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.

[7] M. Karjalainen, V. Valimaki, and T. Tolonen. Plucked-string models: From the Karplus-Strong Algorithm to digital waveguides and beyond. *Computer Music Journal*, 22(3):17–32, Oct. 1998.

[8] M. A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, 1991.

[9] 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*, 2(2):329 –344, Apr. 1994.

[10] 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*, 15(2):531 –541, Feb. 2007.

[11] M. Laurson, C. Erkut, V. Valimaki, and M. Kuushankare. Methods for modeling realistic playing in acoustic guitar synthesis. *Computer Music Journal*, 25(3):38–49, Oct. 2001.

[12] 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.

[13] 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, Oct. 2011. WASPAA.

[14] C. O'Shea. Kinect air guitar prototype, Jan. 2012.

[15] H. Penttinen and V. Valimaki. Time-domain approach to estimating the plucking point of guitar tones obtained with an under-saddle pickup. *Applied Acoustics*, 65:1207–1220, Dec. 2004.

[16] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[17] M. Scholz. Nonlinear PCA toolbox for MATLAB, 2011.

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

[19] J. B. Tenenbaum, V. d. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[20] 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*, 44(5):331–353, May 1996.