

Synthesis of Acoustic Timbres using Principal Component Analysis

Robert G. Laughlin, School of Computing Science
Barry D. Truax, Department of Communication
Brian V. Funt, School of Computing Science

Simon Fraser University
Burnaby, British Columbia
Canada V5A 1S6

Abstract

We have developed an alternate method of representing harmonic amplitude envelopes of musical instrument sounds using principal component analysis. Statistical analysis reveals considerable correlation between the harmonic amplitude values at different time positions in the envelopes. This correlation is exploited in order to reduce the dimensionality of envelope specification. It was found that two or three parameters provide a reasonable approximation to the different harmonic envelope curves present in musical instrument sounds. The representation is suited for the development of high-level control mechanisms for manipulating the timbre of resynthesized harmonic sounds.

Introduction

Our research addresses some problems inherent in the digital generation and manipulation of musical sounds that mimic the natural timbres produced by acoustic, harmonic instruments. The major problem addressed is the large amount of information required to represent the time-varying spectra of harmonic sounds and the difficulty in interacting with this information to alter timbral properties.

Since the pioneering work of Risset and Mathews [11], considerable work has been done on the *Analysis and Resynthesis* method of generating acoustic-instrument timbre [12]. Harmonic sounds are first analyzed to extract the time-varying amplitude envelopes of the harmonic components as well as their time-varying frequency fluctuations. This information is then used to resynthesize sound using an additive synthesis model. To facilitate the manipulation of the timbral properties of the resynthesized sounds, the information extracted from the analysis can be simplified in order to reduce the amount of data involved in sound specification. Line-segment approximation to harmonic amplitude envelopes is one such technique that has been widely used [4, 6, 7, 8, 11]. Further data reduction is possible using line-segment envelopes as a base by taking advantage of interharmonic correlations [1, 5, 13].

Approximating harmonic amplitude envelopes with line segments has two major problems, however. The first is that the extraction of line-segments from the original analyzed curves is often tedious and heuristic (see [15] for an attempt to automate the process). The second and more serious problem is that the representation is not standardized. Since the optimal number of breakpoints and their positioning depends on the nature of the amplitude curve under consideration, it is difficult to compare line-segment-approximated envelopes over different harmonics within a sound or over sounds as a whole. This problem is compounded by the diversity of envelope curves that can result from even one instrument. The set of harmonic amplitude envelope curves characterizing an instrument's sound will be significantly altered by changes in note register, note intensity, room acoustics and player interaction with the instrument. While line-segment approximations have been very useful in studies on the psychoacoustics of timbre [2, 3, 4, 17] they are less well suited for the development of powerful and flexible synthesis tools based on the Analysis and Resynthesis method.

Principal component analysis is a multivariate statistical technique useful in reducing the dimensionality of a large amount of data. It has been used to reduce vowel spectra information [10, 14, 18] and a related technique—multidimensional scaling—has been used to reduce the dimensionality of timbral percepts [2, 3, 4, 9, 16, 17]. Grey [2, 3, 4] and Wessel [17] also correlate the low-dimensional *timbre spaces* resulting from multidimensional scaling with physical correlates of the sound signal. In a 2-dimensional timbre space, the

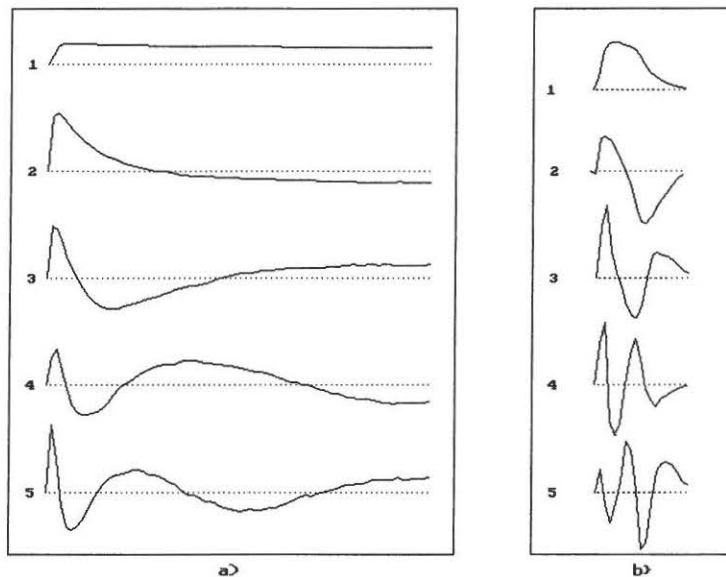


Figure 1: a). Principal component bases for sounds of 2.7 seconds duration. The first 5 principal component bases extracted from the harmonic amplitude envelopes of 24 sounds each of tenor saxophone, clarinet, and trombone, as well as 49 piano sounds. The woodwind sounds have no natural decay. b). Principal component bases for sounds of .65 seconds duration. These bases were extracted from the the harmonic amplitude envelopes of 24 sounds each of trombone, clarinet, and flute. All sounds have a natural decay.

first dimension is related to the *spectral energy distribution* of the sound and the second dimension is related to *temporal* properties of the signal, for example, the quality of the “bite” of the attack [17].

Envelope Representation

Given the low-dimensional nature of the processes involved in timbre perception and the correlation of timbre space dimensions with simple physical properties of the associated sounds, we propose a representation of harmonic amplitude envelopes based on a principal component analysis of a large number of envelopes extracted from a frequency analysis of actual sounds. Since the set of harmonic amplitude envelopes associated with a sound plays a major role in characterizing its timbre, reducing the dimensionality of the envelopes should aid in the ability to manipulate the timbre of sounds at a higher level.

The principal component analysis produces a set of orthogonal basis vectors from a large number of harmonic amplitude envelopes that are extracted from a time-varying frequency analysis of the range of sounds under consideration. The analysis also yields basis-vector weights for each amplitude envelope included in the analysis. Approximations to the original envelopes can be obtained by summing the first few weighted basis vectors and the sounds can be reconstructed from the approximated envelopes using additive synthesis. The envelopes can be manipulated prior to resynthesis by altering the weights associated with each basis vector used in the reconstruction.

Implementation

A rudimentary implementation of the method has been undertaken for 280 sounds from 7 instruments; trombone, clarinet, flute, tenor saxophone, piano, steel-string guitar, and nylon-string guitar. Sounds of 2.7 seconds duration (208) or .65 seconds duration (72) were digitized at a sampling rate of 28,000 samples per second with 8-bit sample resolution. The harmonic amplitude envelopes for the first 20 harmonics (or up to the Nyquist limit) of

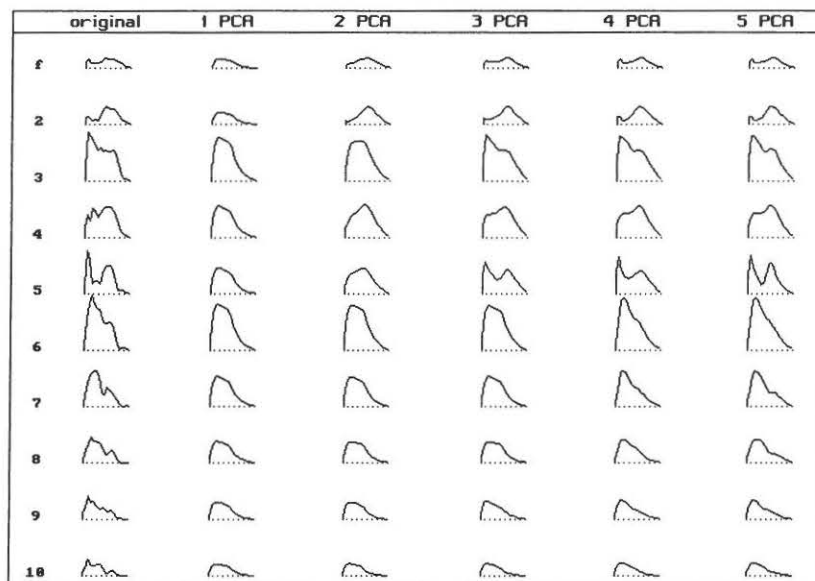


Figure 2: Envelope reconstruction of a trombone B note. The original envelopes for the first 10 harmonics of a short trombone B (123 Hz) sound are on the left of the above graph. The envelopes reconstructed with 1, 2, 3, 4, and 5 principal component (PCA) basis-vector approximations for the same sound are on the right.

each sound were extracted with a 1024-sample Fast Fourier Transform yielding envelope information at 36 ms intervals.

A principal component analysis was performed on various subsets of the envelopes. The analysis yields a statistical “goodness of fit” (variance-accounted-for) measure for the successive approximations resulting from envelope reconstructions with the first few basis vectors. Including disparate instruments (such as piano and clarinet) in the same analysis did not significantly alter the variance-accounted-for which was approximately 90% for the first component, and 97% for the first and second components combined. The basis vectors for some of the 2.7 and .65 second sounds are illustrated in figures 1a and 1b respectively. Figure 2 illustrates the various levels of envelope approximation resulting from the reconstruction of a short trombone sound with the basis vectors in figure 1b.

Evaluation

In a subjective evaluation, sounds resynthesized with as few as 2-basis approximations to the original envelope curves were perceptually similar to sounds resynthesized with the original curves. All resynthesized sounds (including those resynthesized with the original envelopes) were distinguishable to some degree from the original digitized sounds. This is most likely due to the lack of frequency-fluctuation information and inharmonic attack information in the resynthesized sounds. Sounds resynthesized with 1 basis maintained some of the timbral characteristics of the original instrument but tended to sound dull and uninteresting. Since good results have been obtained with 2-basis approximations, harmonic amplitude envelopes can be characterized with only two scalar weights. Additional basis vectors can be used to reconstruct more accurate approximations to the harmonic amplitude envelopes since they tend to capture progressively more of the microstructure of the envelopes; however, the perceptual impact of including more than three basis vectors in the reconstruction is negligible.

A simple perceptual interpretation can also be assigned to the first two basis vectors which aids the user in manipulating timbre. The first basis-vector weight represents the overall contribution of each harmonic to the sound since the first principal component is a simple average of all the amplitude curves as they vary over time. The first basis vector

is akin to a Helmholtz (steady-state) approximation to timbral sounds—except that gross dynamic features are also included. For example, a graph of the first basis-vector weights for a clarinet sound, over the first 20 harmonics, clearly illustrates the odd-harmonic emphasis of clarinet sounds. This correspondence of the first-basis vector to the overall harmonic amplitude levels indicates that the set of first-basis weights—considered as a whole over all the harmonics of a sound—yields a rough measure of the *spectral energy distribution* of the sound. While the second and succeeding bases will alter the envelope *shapes*, their gross contribution to the overall amplitude is much smaller (on the order of 15%). In addition, higher order bases both add and subtract amplitude components since the bases are positive and negative at different time positions in the vectors.

The second basis vector is largely responsible for shaping the attack/decay characteristics of the harmonic components. A positive weight for the second basis increases the attack portion of the harmonic and subtracts amplitude from the latter part of the harmonic. A negative weight for the second basis alters the first basis contribution to produce a slower, less intense attack that builds towards the end of the harmonic component. These interpretations of the roles played by the first two bases correspond to the interpretations of the first two timbre space dimensions in the work of Wessel [17].

High-level Synthesis Tools

In addition to the compact, standardized representation of timbral sounds, the basis-vector weights can be manipulated in an intuitive fashion to produce subtle timbral variations that are constrained in a useful way. Interpolation between timbres is made possible by the common-basis representation underlying all harmonic amplitude envelopes. At the lowest level, the basis weights associated with a sound can be manipulated either individually or as a group to affect timbral changes. The most useful way to display the weight information is in a separate graphical display for the set of weights associated with each basis. We are currently developing a graphical user interface on DSP equipped microcomputers suitable for this level of sound manipulation¹. Constraints (or suggested constraints) can be displayed along with the current weight values. The perceptual interpretation assigned to the bases helps in selecting weight values. At a higher level, the computer itself can be used to select or interpolate between sounds since the underlying representation for all envelopes is the same. In this case, interaction with the sound construction program would be at an abstract level without requiring any direct manipulation of basis weights.

Conclusions

A principal component analysis of the harmonic amplitude envelopes of 280 musical instrument sounds was quite successful at reducing the dimensionality of the envelope curves. The information contained in each curve is automatically extracted and represented by as few as 2 parameters without the loss of perceptually important features of the resulting sound. The basis vectors acquired in the analysis serve as a standardized representation in order to reconstruct and manipulate timbral sounds. The compact representation allows the cataloging of a large number of subtle influences on timbre such as instrument, note register, note intensity, and various player parameters. Sounds not included in the original analysis can be expressed in terms of the representation by taking the dot product of the component envelope curves with each basis vector to obtain the corresponding basis weights. Hence, the analysis is separate from the resynthesis and needs to be performed only once.

In addition to the simple reconstruction of analyzed sounds, the basis vectors and weights can be used as the underlying representation on which to build high-level interactive synthesis tools for the creation of new sounds and the alteration of existing ones. Constraints on the timbral range of sounds produced can be incorporated into the synthesis mechanisms. The set of weights associated with the first two bases have a simple perceptual interpretation which helps to conceptualize and anticipate the effect of altering them. For this reason, the method may be also be useful as a research tool in the study of the psychoacoustics of timbre.

¹Sound resolution is being increased to 16 bit at a 44.1 kHz sampling rate.

References

- [1] Gérard R. Charbonneau. Timbre and the perceptual effects of three types of data reduction. In Curtis Roads, editor, *The Music Machine: Selected Readings from Computer Music Journal*, pages 521–530. The MIT Press, Cambridge, Massachusetts, 1989.
- [2] John M. Grey. Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5):1270–1277, May 1977.
- [3] John M. Grey and John W. Gordon. Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, 63(5):1493–1500, May 1978.
- [4] John M. Grey and James A. Moorer. Perceptual evaluations of synthesized musical instrument tones. *The Journal of the Acoustical Society of America*, 62(2):454–462, August 1977.
- [5] Piotr Kleczkowski. Group additive synthesis. *Computer Music Journal*, 13(1):12–20, Spring 1989.
- [6] James A. Moorer, John M. Grey, and J. M. Snell. Lexicon of analyzed tones part 1: A violin tone. *Computer Music Journal*, 1(2):39–45, 1977.
- [7] James A. Moorer, John M. Grey, and John Strawn. Lexicon of analyzed tones part 2: Clarinet and oboe tones. *Computer Music Journal*, 1(3):12–29, 1977.
- [8] James A. Moorer, John M. Grey, and John Strawn. Lexicon of analyzed tones part 3: The trumpet. *Computer Music Journal*, 2(2):23–31, 1978.
- [9] R. Plomp. Timbre as a multidimensional attribute of complex tones. In R. Plomp and G. F. Smoorenburg, editors, *Frequency Analysis and Periodicity Detection in Hearing*. A. W. Sijthoff, Leiden, 1970.
- [10] R. Plomp, L. C. W. Pols, and J. P. van de Geer. Dimensional analysis of vowel spectra. *The Journal of the Acoustical Society of America*, 41(3):707–712, 1967.
- [11] Jean-Claude Risset and Max V. Mathews. Analysis of musical-instrument tones. *Physics Today*, 22(2):23–30, February 1969.
- [12] Jean-Claude Risset and David L. Wessel. Exploration of timbre by analysis and synthesis. In Diana Deutsch, editor, *The Psychology of Music*, Series in Cognition and Perception, chapter 2, pages 25–58. Academic Press, New York, 1982.
- [13] Keith W. Schindler. Dynamic timbre control for real-time digital synthesis. *Computer Music Journal*, 8(1):28–42, Spring 1984.
- [14] C. L. Searle. Speech perception from an auditory and visual viewpoint. *Canadian Journal of Psychology*, 36(3):402–419, 1982.
- [15] John Strawn. Approximation and syntactic analysis of amplitude and frequency functions for digital sound synthesis. In Curtis Roads, editor, *The Music Machine: Selected Readings from Computer Music Journal*, pages 671–692. The MIT Press, Cambridge, Massachusetts, 1989. reprinted from *Computer Music Journal*, Vol. 4, No. 3, Fall 1980.
- [16] Lage Wedin and Gunnar Goude. Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology*, 13:228–240, 1972.
- [17] David L. Wessel. Timbre space as a musical control structure. *Computer Music Journal*, 3(2):45–52, 1979.
- [18] Stephen A. Zahorian and Martin Rothenberg. Principal-components analysis for low-redundancy encoding of speech spectra. *The Journal of the Acoustical Society of America*, 69(3):832–845, March 1981.