Protoyping and Interpolation of Multiple Musical Timbres Using Principal Component-Based Synthesis

Gregory J. Sandell and William L. Martens
Center for New Music and Audio Technologies, U.C. Berkeley
1750 Arch St, Berkeley CA 94709
sandell@cnmat.berkeley.edu

ABSTRACT

Principal Component Analysis (PCA) has been shown to be an effective technique for representing the complexity of time-variant musical tones efficiently. We explore here a novel application of PCA to Phase Vocoder Analysis (PVA) datasets that yields both common and distinct patterns of spectral evolution for groups of instrument tones. A prototypical timbre is synthesized for the group from the spectrotemporal patterns common to the group's members. Interpolation is accomplished by applying to the prototype the spectrotemporal patterns that distinguish between members. We also introduce a variable-duration temporal partitioning technique for preprocessing the PVA datasets that reduces their size by roughly a factor of 8 while retaining high temporal resolution for perceptually critical portions (attack and decay). This downsampling improves the sensitivity of the PCA to important details of the PVA data, and produces more natural-sounding attack and decay transients when resynthesizing at a variety of durations.

1. INTRODUCTION

Many researchers have explored data reduction techniques for efficient reduction of additive synthesis data sets without loss of perceptually important detail (see Laughlin *et al*, 1990, for a list of sources). Principal Component Analysis (along with the related Karhunen-Loève Transform) appears to be an effective candidate that offers additional insight into high-level features of sounds (Laughlin *et al*, 1990; Stapleton and Bass, 1988), suggests possibilities for realtime control (Weeks *et al*, 1991), and performance gesture identification (Stautner, 1983).

In this paper, we report on three new developments that improve Phase Vocoder Analysis (PVA) data reduction both qualitatively and quantitatively. The first development is a variable-duration temporal partitioning (VDTP) of the PVA datasets that emphasizes attack and decay transients. The second development is the application of PCA to spectral profiles across time, in contrast to the approach of Laughlin *et al* (1990) which applied PCA to temporal amplitude envelopes across partials. The third development addresses the question of how interpolations between various timbres may be obtained, using PCA in combination with other multivariate statistical techniques.

2. DOWNSAMPLING

An important reason for our downsampling the PVA datasets before submitting them to PCA is to emphasize the most perceptually critical portions. While the steady state portions of instrument tones can survive a significant downsampling in temporal precision, the attack and decay transients cannot. Therefore, we have broken the timecourse of each tone into 200 partitions of variable duration. For our PVA datasets containing a total of 1200 to 1800 analysis time frames, roughly 80 of the 200 partitions are dedicated to capturing attack transients at the original PVA frame rate. During the relatively slowly-varying bulk of the tone, partition durations gradually increase to a maximum of roughly 24 frames, and then decrease toward the end of the tone in order to capture the decay at the original frame rate (2.14 ms per frame). Fig. 1 shows a typical distribution of partition durations. In 1a the number of PVA time frames subsumed in each of the 200 partitions are shown; in 1b, the varying duration of the partitions is shown with respect to the original 1508 PVA time frames.

In order to shift a dataset back to its original temporal resolution, a spline is used on each harmonic amplitude envelope. Nearly half of the partition values are unaffected by the resampling since they correspond to single analysis frames. The detail that was lost in the process of downsampling and resampling had virtually no audible consequence for any of the sounds we investigated (wind instument tones from the McGill University Master Samples and ProSonus CDs). Although steady-state portions were somewhat smoother than the original analysis as a consequence of the splining technique, this did not seem to injure the natural quality of the tone. Furthermore, the whole process runs without any user intervention, in contrast to the segmentation required for quality Karhunen-Loève synthesis (Stapleton and Bass, 1988).

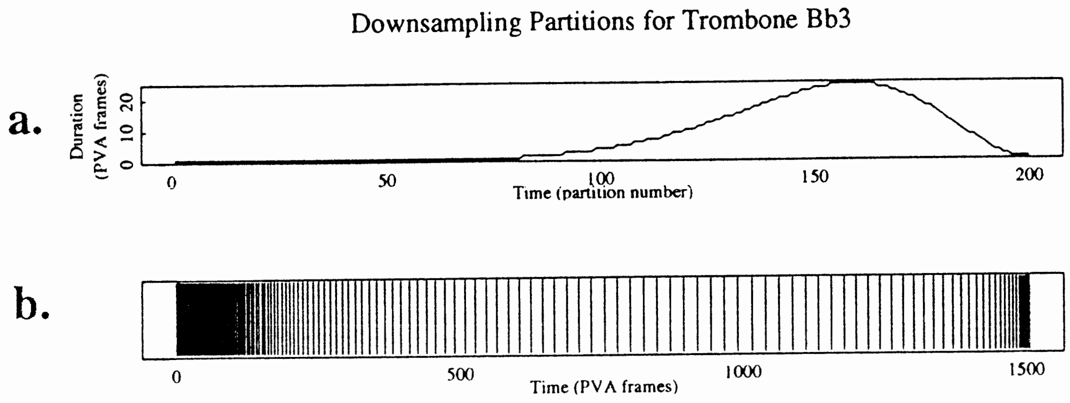

Figure 1

Since attacks and decays are not significantly compressed in downsampling, they sound natural even when a 4-second tone is shortened to 0.5 seconds and vice-versa. This complements the popular time expansion and contraction effects afforded by the phase vocoder that rescale time without changing pitch or spectral energy distribution. And since we regenerate each harmonic amplitude envelope independently, spectral evolution continues throughout the course of the tone, albeit quite slowly during the steady state portion of a time-expanded tone.

One of the major advantages of VDTP is that it maps tones of unequal duration into a common-resolution non-linear time domain. In effect, we have overcome the obstacle of finding the constant number of temporal variates required for analyzing multiple tones in a single PCA, making possible a novel application of PCA to PVA datasets. As the next section of our paper explains, and in contrast to Laughlin *et al* (1990),we submit the transpose of the matrix of PVA data to PCA, a technique we call *spectral PCA*.

3. SPECTRAL PRINCIPAL COMPONENTS ANALYSIS

We describe PCA here informally as it relates to the context of additive synthesis data reduction. For more formal detail, the reader is referred to Harris (1985) and Martens (1987). Although our exposition here refers to the manipulation of amplitude information only, our work includes equal treatment of both the amplitude and frequency information from PVA datasets.

PVA datasets of instrument timbres contain inherent redundancy since the temporal amplitude envelopes of partials in natural instrument sounds are correlated, as in Fig. 2. This redundancy can be reduced to just a few dimensions of orthogonal basis vectors; for example, most of the activity in harmonics 4-12, with their early sharp peak, can be captured by one basis vector, while the slower-rising, flat-peaked shape of harmonics 1-3 can be captured by another. Successive principal components (PCs) add features to account for the idiosyncratic features of one or more other harmonics (such as the late "bumps" in partials 11 and 13) and/or as a corrective term for features that must be removed from certain harmonic amplitude envelopes. In this approach, the one adopted by Laughlin *et al* (1990), a PVA dataset is input to PCA as a series of temporal envelopes, one for each harmonic; consequently, the basis vectors themselves resemble temporal envelopes. We refer to this approach as *temporal PCA*. An alternative view of a PVA dataset is as a series of spectral envelopes for each time frame, as shown in Fig. 3 (here the entire tone is shown in only 20 frames). Each of the four basis vectors shown (top row) resembles a spectral envelope, and each successive component adds more of the detail that distinguishes the spectral envelope at each point in *time*. We refer to this as *spectral PCA*.

To make the distinction between these two methods clear, we must examine the orientation of the matrix input to PCA: If the downsampled PVA data matrix is organized with rows corresponding to frequency values, and columns corresponding to time values, then we have spectral PCA. This analysis capitalizes on correlations between the columns of spectral envelopes. If the PVA data matrix is transposed, we have temporal PCA that capitalizes on correlations between the columns of temporal envelopes. Note that without the VDTP downsampling, spectral PCA on the original PVA data would require the computation of a huge covariance matrix (1508 by 1508 for the trombone example shown in Fig. 1). The rank of the covariance matrix for temporal PCA is determined by the number of harmonics (22 in our examples).

Both methods result in an efficient reduction of the data by the determination of orthogonal vectors in the dataspace. An exact recovery of the original dataset is guaranteed by using all PCs. However, when using only the 3 PCs we found necessary to capture the perceptually relevant variance, the result was data reduction greater than a factor of 67 in our examples. Moreover, PCA can reveal high-level features that may be perceptually relevant and/or useful for musical manipulation, but which are not apparent by viewing the raw data.

We prefer spectral PCA to temporal PCA for our exploration of timbre for several reasons. As a rule, PCA is an efficient data reduction method only when the input variables are correlated. In natural sounds, the spectral envelope changes smoothly through most of the tone's duration, meaning that there is a great deal of correlation between neighboring time frames. Temporal envelopes, on the other hand, are often less highly correlated (e.g., flute and violin). In fact, all the tones employed in our study showed a higher average correlation of spectra over time frames than correlation of temporal envelopes over harmonics. For example, the resynthesized spectra shown in Fig. 3 resembles the original data in significant detail by the 3rd PC, whereas a resynthesis with temporal PCA (not shown) needed at least 6 PCs to reach the same degree of detail. Admittedly, temporal PCA might be more efficient for tones with a great amount of spectrotemporal flux; but for the type of tones we investigated, we found spectral PCA most efficient.

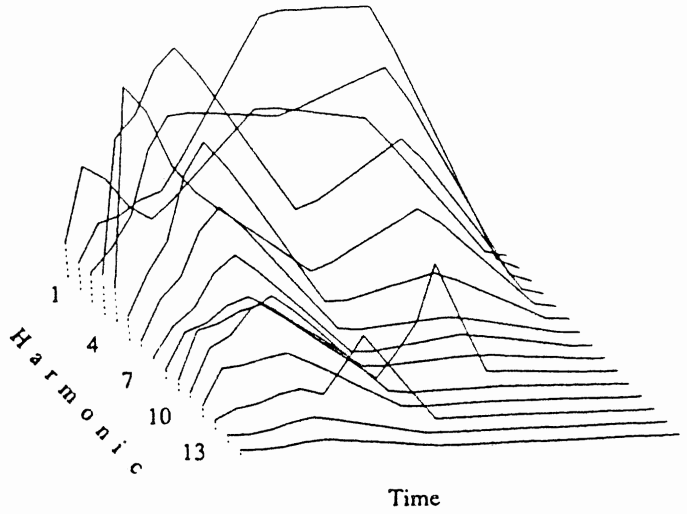

Figure 2

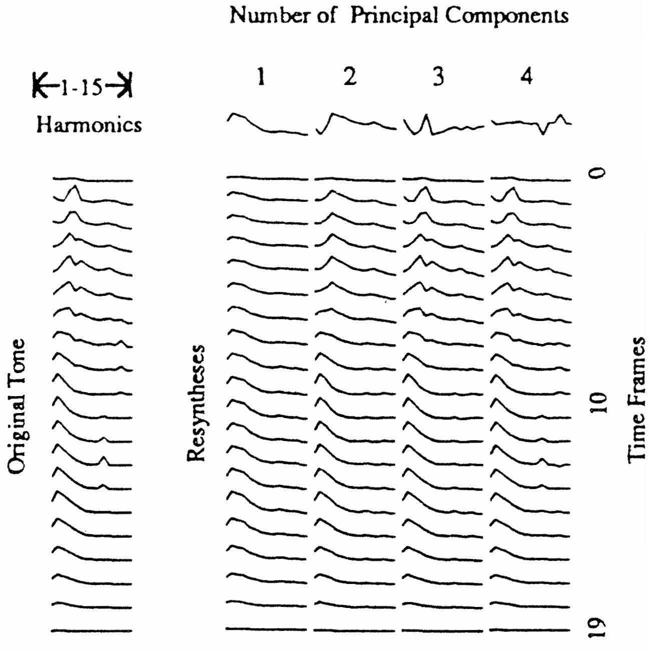

Figure 3

Furthermore, spectral PCA is more space efficient when analyzing multiple instrument tones. This is because the basis vectors are usually longer in temporal as compared to spectral PCA. Though the weighting functions are perforce longer in spectral PCA, these are common to all instruments, so as the number of instruments grows large, the amount of data required for temporal vs. spectra PCA depends only upon the length of the basis vectors. In an example using 3 PCs (and PVA datasets of 22 frequency values by 200 time values), we found data-reduction roughly 10 times greater using spectral PCA. This obviously has important ramifications for synthesis using realtime hardware since spectral PCA can store information on spectral evolution for many more instruments in the same amount of memory. Experiments for realtime synthesis on the Silicon Graphics *Indigo* are underway using techniques described in Freed (1992) and Rodet and Depalle (1992).

From a more theoretical perspective, we posit that a characterization of spectral differences over time may be more useful than information on the envelope differences across harmonics. For example, consider a PCA of several tones from the same instrument (different pitches, articulations). Since spectral properties are more likely to be invariant across multiple durations and articulations of an instrument than temporal properties, a single spectral PCA may offer more insight into an instrument's general timbral identity.

4. PROTOTYPING AND INTERPOLATION

We have described a mechanism for packaging tones of unequal duration into downsampled PVA data matrices of 22 frequency values by 200 time values. We have also shown that these data can be simplified further through the application of spectral PCA. Now we show the advantage of breaking that spectral PCA into two parts, employing methods from Multivariate Analysis of Variance (MANOVA; see Harris, 1985). The first part is the generation of a prototypical timbre from a set of timbres using a PCA that finds weighting functions only for what the set of timbres has in common. The second part is the interpolation between those timbres using a PCA that finds weighting functions only for what distinguishes between the timbres. The prototype weighting functions were generated by finding the eigenvectors of the pooled within-groups sums of squares and crossproducts (SSCP) matrix. The interpolation weighting functions were based upon the between-groups SSCP matrix. The within-group and between-group SSCP matrices sum to equal the total SSCP matrix for all the PVA datasets submitted, hence the information for complete reconstruction of the original datasets is not lost in these operations.

By isolating what all the tones have in common, we can generate a prototypical tone that captures none of the features of any tone in particular. When several different horns were analyzed, the resulting prototype sounded like a bland, generic horn sound. This result is perhaps not musically useful, but if we subtract the prototype matrix from all the individual instrument matrices, the prototype becomes the origin of a multidimensional coordinate system within which each instrument is located at a unique point relative to that origin. This has the musically useful result of setting up a control structure for timbral interpolation. To reduce the dimensionality of the interpolation coordinate system, the deviation matrices were scored on the interpolation weighting functions. Using only the first 3 PCs for the deviation scores creates a coordinate system that can be easily explored.

The primary source of information for synthesis is in the PCs for deviation scores. Once the prototype is generated, it becomes a "center of gravity" for the space within which interpolation takes place. Note that a prototypical timbre could have been synthesized from a simple average of all the PVA data matrices in the set of timbres, but we found that this result had objectionable idiosyncrasies that did not appear in the more idealized PC-based prototype. Since the interpolation space is defined by 3 PCs, there are multiple paths between each of the analyzed timbres. We can take the shortest path between two timbres, or follow a piecewise linear path that changes values on one dimension at a time.

5. CONCLUSION

We have reduced and regularized the data of phase vocoder analysis through VDTP downsampling (reduction factor of roughly 8), and then reduced it further via spectral PCA. The fixed memory-size cost of PC-based synthesis is 200 points per PC weighting function, regardless of the number of timbres represented. If 3 PCs are employed, then 200 PVA frames can reduced to 3 basis vectors (a reduction factor of 67 for each timbre). Ignoring the fixed-size weighting function matrix, the combinination of these two procedures gives a reduction factor of more than 500 with respect to the original PVA data. We note that the data reduction provided by VDTP downsampling is not nearly as great as that obtained by fitting line-segment approximations to each amplitude envelope (see Grey, 1975), but such approximations are not suitable for PCA, and do not evince many of the desireable features of VDTP downsampling.

There are several advantages offered by these methods for compositional exploitation of timbre. Spectral PCA allows us to keep spectral attributes of an analyzed timbre invariant while exploring creative temporal manipulations. Furthermore, the potential for timbre interpolation as described in Wessel *et al* (1987) and Grey (1975) is offered by our method of representing multiple tones.

6. ACKNOWLEDGMENTS

James Beauchamp's PVA software, pvan and monan (see Beauchamp, 1989), was invaluable to us. The first author wishes to thank the Acoustical Society of America for support via the F.V. Hunt Post-doctoral Research Fellowship in Acoustics.

7. REFERENCES

J. Beauchamp, "The Computer Music Project at the University of Illinois at Urbana-Champaign," *1989 ICMC Proc.*, 21-24, 1992.

A. Freed, "New tools for rapid prototyping of musical sound synthesis algorithms and control strategies," *1992 ICMC Proc.*, 1992.

J.M. Grey, *An Exploration of Musical Timbre*. PhD thesis, Stanford University, 1975.

R.J. Harris, *A Primer of Multivariate Statistics*. 2nd Ed., New York: Academic Press, 1985.

R.G. Laughlin, B.D. Truax and B.V. Funt , "Synthesis of acoustic timbres using principal components analysis," *1990 ICMC Proc.*, 95-99, 1990.

W. Martens, "Principal components analysis and resynthesis of spectral cues to perceived direction," *1987 ICMC Proc.*, 274-281, 1987.

X. Rodet and P. Depalle, "Spectral envelopes and inverse FFT synthesis," *Proc. of the AES 93rd Convention*, 1992.

J.C. Stapleton and S.C. Bass , "Synthesis of musical tones based on the Karhunen-Loève Transform," *IEEE ASSP* 36/3, 305-319, 1988.

J.P. Stautner, *Analysis and Synthesis of Music Using the Auditory Transform*, Master's thesis, MIT EECS Dept., Cambridge, MA., 1983.

W.B. Weeks, W.A. Schloss and R.L. Kirlin, "Implementation of the KL Synthesis Algorithm under real-time control," *1991 ICMC Proc.*, 360-363, 1991.

D. Wessel, D. Bristow and Z. Settel, "Control of phrasing and articulation in synthesis," *1987 ICMC Proc.*, 108-116, 1987.