Methods for Multiple Wavetable Synthesis of Musical Instrument Tones*

ANDREW HORNER, JAMES BEAUCHAMP, *AES Fellow,* **AND LIPPOLD HAKEN****

University of Illinois at Champaign-Urbana, IL 61801, USA

Spectrum matching of musical instrument tones is a fundamental problem in computer music. Two methods are presented for determining near-optimal parameters for the synthesis of harmonic musical instrument or voice sounds using the addition of several fixed wavetables with time-varying weights. The overall objective is to find wavetable spectra and associated amplitude envelopes which together provide a close fit to an original time-varying spectrum. Techniques used for determining the wavetable spectra include a genetic algorithm (GA) and principal components analysis (PCA). In one study a GA was used to select spectra from the original signal at various time points. In another study PCA was used to obtain a set of orthogonal basis spectra for the wavetables. In both cases, least-squares solution is utilized to determine the associated amplitude envelopes. Both methods provide solutions which converge gracefully to the original as the number of tables is increased, but three to five wavetables frequently yield a good replica of the original sound. For the three instruments we analyzed, a trumpet, a guitar, and a tenor voice, the GA method seemed to offer the best results, especially when less than four wavetables were used. Comparative results using the methods are discussed and illustrated.

0 INTRODUCTION

Matching synthesis of musical instrument tones is a fundamental problem in computer music. Generally, for a particular synthesis model matching begins with a time-variant spectral analysis of the original sound. Next, the model synthesis parameters which produce a "best fit" to the analysis data are determined. Finally, resynthesis of the sound is performed using the matched parameters. These steps are shown in Fig. 1.

Synthesis models generally fall into one of three categories: 1) time-variant filter synthesis, 2) nonlinear distortion synthesis, and 3) fixed-waveform additive synthesis. Previous matching methods are briefly reviewed with respect to these categories and are illustrated in Fig. 2.

Time-variant filter synthesis includes linear predictive coding (LPC) [1], a method of finding time-varying digital filter parameters for matching sounds. Traditionally the input is either a pulse train waveform or white noise, but there is some debate about what is the "best" input signal. LPC has generally been applied more successfully to speech sounds than to musical tones. It has been very successful for ordinary speech and "singing speech" [2], and works well with variable fundamental frequency, since pitch detection is usually part of the analysis technique. The method can be persuaded to converge to perfection if a sufficient number of filter stages is included.

Finding parameters which enable nonlinear synthesis methods to match acoustic sounds is inherently difficult due to the complex spectral evolution characteristics inherent with these methods. Frequency modulation (FM) and nonlinear processing (waveshaping) are two synthesis techniques which fall into this category. Estimation of the FM parameters of an acoustic sound has been an elusive problem, despite some attempts made in this direction [3]–[5]. Recently the authors of this paper showed how a genetic algorithm (GA) can be successfully applied to matching a single-modulator, multiple-carrier FM model [6]. The GA was used to select fixed values of the modulation indexes and carrier-to-modulator frequency ratios, while the amplitude of each carrier was determined by least-squares solution.

* Manuscript received 1992 July 29; revised 1993 February 15.

** A. Horner and L. Haken are with the CERL Sound Group; A. Horner and J. Beauchamp are with the Computer Music Project.

Spectral centroid matching [3] for nonlinear processing synthesis leads to a relatively simple method for achieving approximations of some acoustic sounds. This technique works best for spectra which are well characterized by a principal spectrum whose shape is modified according to a well-defined time-varying spectral centroid.

Multiple wavetable synthesis, the subject of this paper, is based on a sum of fixed waveforms or periodic basis functions with time-varying weights. Each waveform can be expressed as a fixed weighted sum of several harmonic sine waves. If the sets of harmonics for the various waveforms are disjoint, the method is termed "group additive synthesis" [7]. More generally, Stapleton and Bass presented a statistical method based on the Karhunen–Loève (KL) transform to determine periodic time-domain basis functions (waveforms) as well as amplitude and phase-control functions to optimally fit acoustic tones [8]. Their optimization was based directly on the time signal, so that spectral analysis was not necessary. The same basis functions were used for several different instruments. Two drawbacks of this method are its computationally expensive matching procedure and the phase cancellation problems which potentially arise when waveform amplitudes vary from their designated values.

Waveform or spectrum interpolation synthesis [9] assumes that a signal may be divided into a series of "target" waveforms. Synthesis proceeds by gradually fading (or interpolating) from one target to the next. (Interpolation between waveforms and interpolating between corresponding spectra are the same only if the phases of the corresponding harmonics of the two spectra are the same.) This method might be thought of as the opposite extreme of group additive synthesis, in that spectra are disjoint in time rather than in frequency. Serra et al. give a method based on linear regression whereby target waveforms are selected based on the assumption of linear ramp interpolation between spectra.

This paper presents two general matching methods for selecting additive wavetable spectra, one based on a GA [1], [11], the other on principal components analysis (PCA) [12]. GAs have been applied to a wide array of problem domains from stack filter design [13] to computer-assisted composition [14]. The GA-based spectrum matching methods presented in this paper find parameters which can be used to perform traditional wavetable synthesis. Our PCA-based technique is related to that of Stapleton and Bass; however, our basis functions are determined in the frequency rather than in the time domain. The PCA approach has been used in various speech applications [15], [16]. For both of our methods, the time-varying weights are determined by least-squares or direct matrix solution.

1 WAVETABLE SYNTHESIS OVERVIEW

Wavetable or fixed-waveform synthesis is an efficient technique for the generation of a particular periodic waveform. Prior to synthesis, one cycle of the waveform is stored in a table. The spectrum of the waveform can be an arbitrary harmonic spectrum, which is specified by the amplitude values of its harmonics. The table entries are given by

$$\text{table}_i = \sum_{k=1}^{N_{\text{hars}}} a_k \sin\left(\frac{2\pi ki}{\text{table length}} + \phi_k\right) \qquad (1)$$

where $1 \leq i \leq$ table length, and a_k and ϕ_k are the

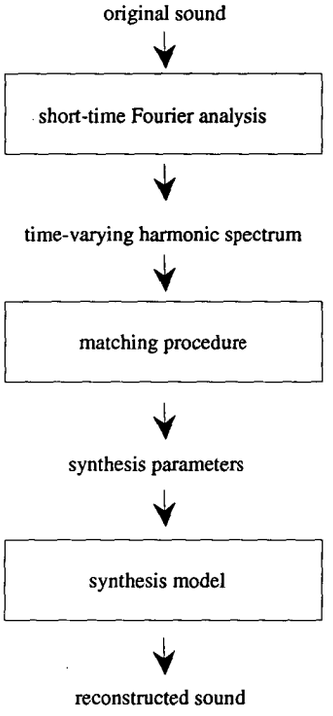


Fig. 1. Wavetable matching analysis/synthesis overview.

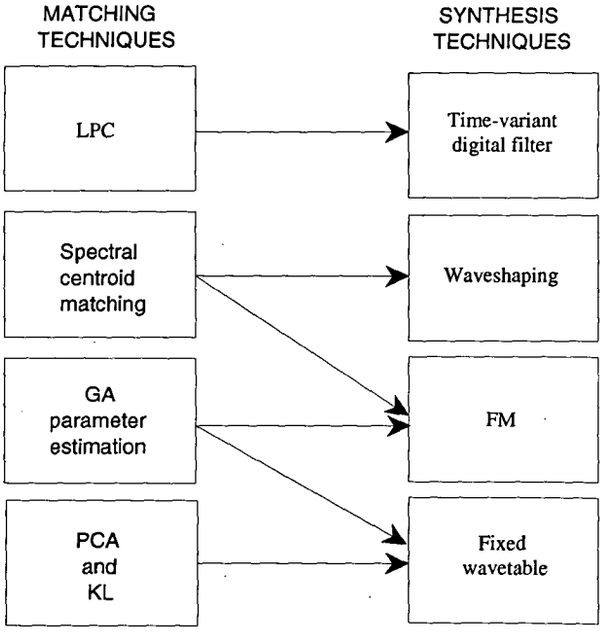


Fig. 2. Matching and synthesis models.

amplitude and phase of the kth partial, and N_{hars} is the number of harmonics needed to represent the signal. The phases ϕ_k are generally not audibly important, and are often simply set to 0 or arbitrary values. The spectrum produced by a particular set of a_k values will be referred to as the wavetable's associated basis spectrum.

To generate samples during synthesis, table lookup is performed for the desired number of samples. Initially the table is indexed at its first entry. Subsequent lookups increment the index by the sample increment and read the sample at the new index point. The sample increment is given by

$$\text{sample increment} = f_1 * \frac{\text{table length}}{\text{sampling rate}} \quad (2)$$

where f_1 is the desired fundamental frequency of the sound. Note that f_1 can be fixed or time varying. Fig. 3 shows the standard symbolic notation for a simple wavetable instrument.

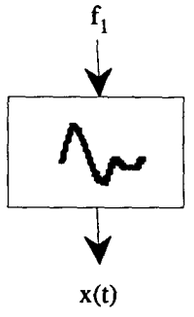


Fig. 3. Simple wavetable model.

The sample increment will generally not be an exact integer. The table index value may be truncated or rounded, or, alternatively, interpolation may be used to improve lookup accuracy. Signal-to-noise considerations generally determine the approach used [17], [18].

Further control can be gained by using multiple weighted wavetables in the synthesis model, as shown in Fig. 4. The time-varying weights on the tables allow them to be cross-faded and generally mixed with one another in various ways. Note that the phases of the corresponding harmonics of multiple wavetables must be the same to avoid inadvertent phase cancellation.

The principal advantage of wavetable synthesis is its efficiency. For each wavetable sample, the main steps are to compute the sample increment (and then only if the fundamental is time varying), perform the waveform table lookup, look up the weights from the envelope tables, and postmultiply the waveforms by the table weights. In terms of storage, only one period of each waveform is needed plus its associated table weights, a relatively inexpensive requirement.

A disadvantage of the technique stems from the fact that each wavetable produces a static spectrum, while real sounds produce dynamic spectra. For an arbitrary small set of wavetables, most time-varying spectra cannot be approximated very closely by a linear combination of these wavetables, even if their weights are time varying. Thus the basis spectra must be chosen carefully and their weights appropriately manipulated when synthesizing dynamic spectra.

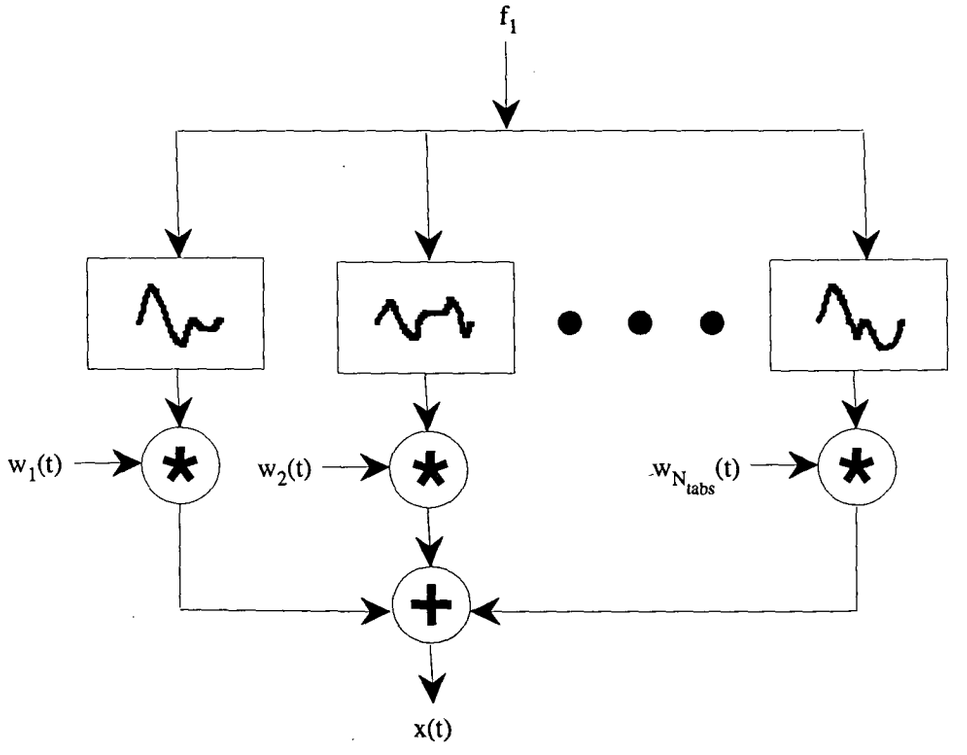


Fig. 4. Wavetable synthesis model.

2 SHORT-TIME SPECTRUM ANALYSIS

Matching of a wavetable-synthesized signal to an original musical signal is facilitated by working in the frequency domain and matching the sound spectra of the original and synthesized signals. The basic assumption we make about any original signal is that it can be represented by a sum of sine waves with time-varying amplitudes and frequencies:

$$y(t) = \sum_{k=1}^{N_{hars}} b_k(t) \sin\left[2\pi \int f_k(t)\, dt + \theta_k\right] \quad (3)$$

where $b_k(t)$ and $f_k(t)$ are the time-varying amplitude and frequency of the kth harmonic in the original signal. Note that in our matching procedure we will ignore the starting phases θ_k since they have no audible effect in most music synthesis situations.

A further restriction that we use in this paper is that the sound is harmonic, so that

$$f_k(t) = kf_1(t) \ . \quad (4)$$

Not all instruments are periodic or quasi-periodic. Thus we would not expect techniques based on such a harmonic restriction to fare as well with less periodic sounds, such as low piano tones, than with other instruments.

We use two different techniques to analyze a sound in order to estimate its harmonic amplitudes and fundamental frequency. Our usual method is a fixed filter bank approach where bandpass filters are centered on the harmonics of an "analysis frequency" which approximates the mean of $f_1(t)$ [3], [19]. In this method the filter outputs give real and imaginary parts, which are converted into amplitudes and phases by the right-triangle solution. The fundamental frequency is then computed from the derivatives (or finite differences) of the phases. However, this method fails if the $f_1(t)$ frequency deviations are so great that upper harmonic frequencies swing on the edges or outside the ranges of the filters, since the harmonic amplitudes would then be seriously in error. Therefore for cases where substantial vibrato or portamento occur, we use an extension of the McAulay–Quatieri (MQ) analysis technique [20], which is capable of tracking severe changes of pitch. By itself, the MQ method computes amplitudes and frequencies on the basis of finding peaks in a spectrum computed by using a fixed-length fast Fourier transform. However, since this method is not inherently restricted to harmonics, harmonic frequencies must be "interpreted" from the data for the peaks. We do this by first estimating the fundamental frequency as a function of time from the MQ spectral data and then by sorting these data into appropriate harmonic bins [21].

3 WAVETABLE SPECTRAL MATCHING

By determining a set of basis spectra and associated amplitude envelopes whose sum best matches the original time-variant spectrum, we attempt to reconstruct the sound using an elaboration of traditional wavetable synthesis. The matching procedure consists of two steps whereby the basis spectra and amplitude envelopes are determined. The principal contribution of this paper is a method for efficient determination of the basis spectra and their envelopes.

As the first step, the user specifies the number of basis spectra to be used in making the match, and the basis spectra are then determined. Two methods for determining basis spectra are given in Section 4.

The second step is to determine the optimum-amplitude envelope (time-varying weight) for each table by straightforward matrix solution. Using the already determined basis spectra and the sequence of discrete-time spectra of the original sound, we form a system of linear equations represented by the matrix equation

$$AW \approx B \ . \quad (5)$$

As Fig. 5 shows, the matrix A contains the wavetable basis spectra stored as a series of columns, with one column for each spectrum; the matrix W contains the unknown amplitude weights, corresponding to time samples of the (as yet undetermined) envelopes for each time frame of the analysis, arranged in a series of columns; and the matrix B contains successive frames of the original discrete-time spectra. This system of equations is of the form

$$\sum_{j=1}^{N_{tabs}} a_{kj}\, w_{j,r} \approx b_{k,r} \quad (6)$$

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,N_{tabs}} \\ a_{2,1} & \cdots & a_{2,N_{tabs}} \\ \vdots & & \\ a_{N_{hars},1} & \cdots & a_{N_{hars},N_{tabs}} \end{bmatrix} \times \begin{bmatrix} w_{1,1} & \cdots & w_{1,N_{frames}} \\ w_{2,1} & \cdots & w_{2,N_{frames}} \\ \vdots & & \\ w_{N_{tabs},1} & \cdots & w_{N_{tabs},N_{frames}} \end{bmatrix} \approx \begin{bmatrix} b_{1,1} & \cdots & b_{1,N_{frames}} \\ b_{2,1} & \cdots & b_{2,N_{frames}} \\ \vdots & & \\ b_{N_{hars},1} & \cdots & b_{N_{hars},N_{frames}} \end{bmatrix}$$

Fig. 5. Matrix representation of Eq. (5).

for $1 \leq k \leq N_{\text{hars}}$ and $1 \leq r \leq N_{\text{frames}}$. In this equation a_{kj} is the time-fixed amplitude of the kth harmonic due to the jth basis spectra, $w_{j,r}$ is the envelope weight for the jth basis spectra at the rth time frame, and $b_{k,r}$ is the amplitude of the kth harmonic of the analysis spectrum at the rth time frame. Note that the duration of the sound under analysis is $t_{\text{dur}} = \Delta t\, N_{\text{frames}}$, where Δt is the duration of a single frame. If the number of basis spectra N_{tabs} is equal to the number of harmonics N_{hars} of the original sound and assuming that the basis spectra are independent, Eqs. (5) and (6) can be solved exactly by direct matrix solution, so a perfect solution, analogous to ordinary sine-wave additive synthesis, results. But what we want is a reduced set of basis spectra. For this case we can determine a best solution in the least-squares sense. This is tantamount to determining the $\{w_{j,r}\}$ that minimize the squared error

$$\sum_{k=1}^{N_{\text{hars}}} \left(\sum_{j=1}^{N_{\text{tabs}}} a_{kj}\, w_{j,r} - b_{k,r} \right)^2 \tag{7}$$

for each time frame r. Note that each time frame is independent and could be solved without consideration of other time points. However, a more efficient solution results when time frames are considered as a group. Indeed, there exist very efficient algorithms for a direct least-squares solution of Eq. (5), such as solution by the use of the normal equations [22]. Thus for a given set of basis spectra, the computation of their amplitude envelopes is a straightforward process.

Specifically, we must solve for W in the symmetric linear system

$$A^{\mathrm{T}}AW = A^{\mathrm{T}}B \tag{8}$$

known as the normal equations, where A^{T} is the transpose of the matrix A. In order to solve for the weights W, $A^{\mathrm{T}}A$ must be nonsingular, and this is true only if the columns of A are linearly independent, that is, the basis spectra are linearly independent of one another. This is a reasonable requirement, since we will naturally wish to find a minimal set of dissimilar basis spectra which efficiently spans the spectral space.

In unusual cases the normal equations will have problems due to inaccuracies resulting from finite precision arithmetic. For instance, information can be lost in forming the normal equations matrix $A^{\mathrm{T}}A$ as well as the right-hand side matrix $A^{\mathrm{T}}B$. In these situations, orthogonalization methods, such as QR factorization [22], can be used instead of the normal equations, since they do not amplify error. However, the improved accuracy afforded by these methods is accompanied by increased computational expense. In practice, when only a few wavetables are used, solution by the normal equations will generally suffice. However, if inaccuracies manifest themselves in the form of unusually large weights, orthogonalization methods should be considered [23].

Back to the first step—how should we determine the basis spectra? Section 4 presents two approaches for solving this problem. One uses a GA to determine the basis spectra. The other is based on PCA.

3.1 The Relative Error Measure

The *relative error* is used to measure the quality of the match between a candidate synthetic signal and the original signal. In the case of the GA approach, the relative error is used as a fitness measure to guide the search for a good solution. We define the relative error as

$$\bar{\varepsilon}_{\text{rel}} = \frac{1}{N_{\text{frames}}} \sum_{i=1}^{N_{\text{frames}}} \left\{ \frac{\sum_{i=1}^{N_{\text{hars}}} [b_k(t_i) - b'_k(t_i)]^2}{\sum_{k=1}^{N_{\text{hars}}} b_k^2(t_i)} \right\}^{1/2} \tag{9}$$

where the t_i are particular selected time values within the duration of the sound being matched, N_{frames} is the number of time values selected, $b_k(t)$ is the kth harmonic amplitude of the original signal, N_{hars} is the number of harmonics, and

$$b'_k(t) = \sum_{j=1}^{N_{\text{tabs}}} w_j(t)\, a_{k,j} \tag{10}$$

is the time-varying amplitude of the kth harmonic of the synthesized signal. Obviously we would expect that the lower the value of this error measure, the better the perceptual match. Our experience so far is that this is generally but not always true. However, lacking a formula which is a good predictor of subjective preference, this is what we are using for the time being.

The computational cost of computing Eq. (9) for a candidate solution is reduced considerably by restricting the time average to a limited number of representative spectra from the sound being matched, rather than using all of the analysis frames. Judicious choices of spectra from the original time-variant spectrum are important for achieving a reasonable wavetable approximation. For example, spectra in the attack portion of a sound are very good choices, since the attack is a perceptually critical and a fast-changing portion of the tone [24]–[26].

In fact, some matches were found to be perceptually better when only a few spectra were used in the error calculation instead of utilizing all of the analysis spectra available. This somewhat surprising result makes sense when one considers that the discrete spectra which most dominate the time average come from the comparatively long sustain portion of the tone. Perceptually important spectra occurring in the brief attack are simply overwhelmed by these spectra. We needed a method to avoid this problem. After considerable experimentation we arrived at a spectrum selection procedure based on picking 50% of our representatives from the "attack"

portion of the tone (defined as the portion before the peak rms occurs), and the other 50% from the remainder of the tone. The specific method which was most successful simply picked the spectra from equally spaced time points in these two regions, as depicted in Fig. 6(a). An alternative approach entailed picking spectra equally spaced in amplitude rather than time, as shown in Fig. 6(b). However, this latter approach fails if the

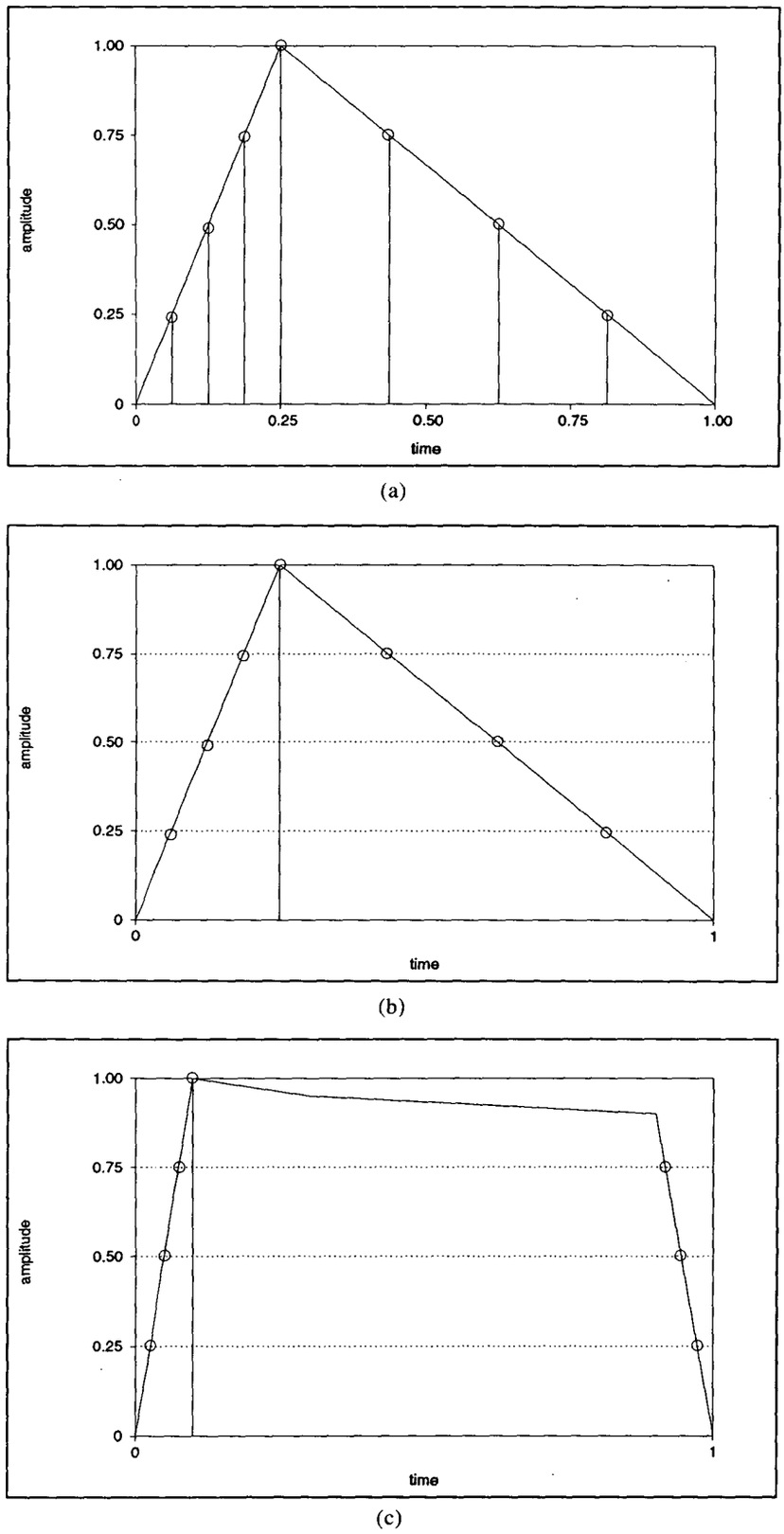


Fig. 6. (a) Selected spectra equally spaced in time on both sides of peak. (b) Selected spectra equally spaced in amplitude on both sides of peak. (c) Problematic case for picking spectra equally spaced in amplitude.

attack and decay are too short. Fig. 6(c) is an example of such a case. The first approach appears to be more robust under these conditions.

4 METHODS FOR DETERMINING BASIS SPECTRA

As depicted in Fig. 7, there are a number of possible approaches for generating basis spectra. In general either one may select basis spectra from the set of short-time spectra found by Fourier analysis of the original sound, or one may generate spectra based on a suitable algorithm.

4.1 GA-Based Selection

As mentioned earlier, we could, without applying any initial criteria on the types of candidate spectra to be considered, simply let the GA routine search for the best relative harmonic amplitudes for each basis spectrum. However, an obvious problem with allowing each harmonic to take on any value between 0 and 1 is that with most sounds, the higher harmonics tend to have relatively small amplitudes. Thus a large portion of the candidate solution space would yield very poor matches. This could be remedied by setting the upper bound for the search range of each harmonic's relative amplitude to be that harmonic's maximum over the duration of the tone. Even so, another, more general, disadvantage with this scheme is that it intrinsically gives rise to a large number of variables, thus defining a huge search space with a relatively small number of usable solutions. The dimension of this space is equal to the product of the number of basis spectra to be determined and the number of harmonics for each spectrum. However, we can speed up the search considerably if we can find methods which reduce the number of variables to only one or two for each basis spectrum.

One simple but very effective approach is to use a GA to pick basis spectra from the sound's own set of discrete-time spectra. This method bears some resemblance to spectral interpolation [9], where spectra picked from different parts of the original sound are simply cross-faded from one to the next as time progresses. In this case, only one parameter per basis spectrum is needed, an index corresponding to the time of the chosen analysis spectrum. Initially the GA considers a population of randomly chosen indexes corresponding to particular basis spectra. Subsequently the GA mixes and matches these choices in an attempt to determine a set of basis spectra which work well over the course of the tone. We have found this GA-index technique to be the most successful we have tested so far, in that the computation time to determine good parameter values is quite low, and the results give the lowest average relative errors as well as the best subjective results.

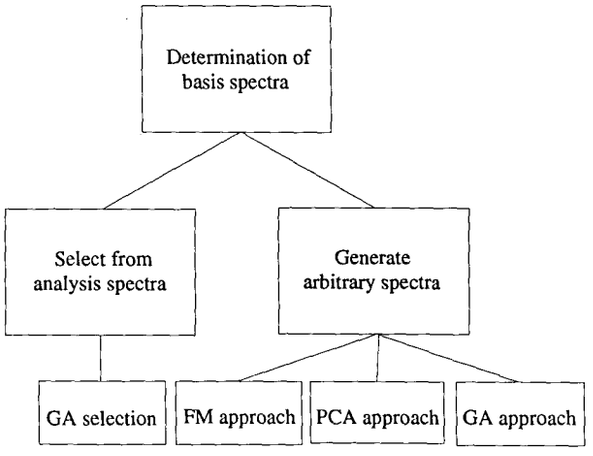


Fig. 7. Hierarchy of basis spectra-generation methods.

Another method, which we explore in a companion paper [6], uses FM basis spectra. Fixed FM spectra are very special cases of wavetable spectra. If a single modulator modulates a carrier and its modulation index is held constant, a static spectrum results. A harmonic spectrum results whenever the carrier-to-modulator frequency ratio is an integer. Using a single modulator and several carriers, each having a different index and carrier-to-modulator ratio, we get a set of basis spectra (one for each carrier) as in the preceding cases. Thus each basis spectrum is characterized by two variables. The GA-FM technique will not be pursued here, but interested readers are invited to consult the companion paper.

4.2 Principal-Components-Based Matching

Basis spectra can be determined by statistical factor analysis procedures, and PCA is one such procedure which offers an elegant solution to the wavetable matching problem. This method has the advantage that the derived basis spectra will be optimal in a statistical sense—they capture the maximum variance of the analyzed tone. Moreover, the basis spectra found are guaranteed to be orthogonal to one another (that is, any one of them may not be expanded as a weighted sum of the others). Finally the PCA method ensures that, for a given number of fixed basis spectra, the time-averaged mean-square error between the original and the matched spectra will be minimal. On the other hand, these basis spectra may have a rather artificial relation to the original tone, since generally no basis spectrum will resemble any of the actual analysis spectra. In any case, the PCA technique offers an interesting alternative to the GA-based techniques, which, in general, we have found to be more successful.

PCA determination of the basis spectra consists of three steps, as illustrated in Fig. 8. Recall that the original analysis spectra are contained in the matrix B of Fig. 5. In the first step we form the covariance matrix C from the original tone's short-time spectra in the matrix B. Each entry of C can be found using the equation

$$c_{k_1,k_2} = \frac{1}{N_{\text{frames}}} \sum_{i=1}^{N_{\text{frames}}} (b_{k_1,i} - \bar{b}_{k_1})(b_{k_2,i} - \bar{b}_{k_2}) \quad (11)$$

where $k_1, k_2 = 1, 2, \ldots, N_{\text{hars}}$, and

$$\bar{b}_k = \frac{1}{N_{\text{frames}}} \sum_{i=1}^{N_{\text{frames}}} b_{k,i} \tag{12}$$

with $b_{k,i} = b_k(t_i)$.

The covariance matrix contains the variance of each harmonic's amplitude on its respective diagonal. The nondiagonal elements are the covariances between the respective harmonics, which is simply the mean of their cross products as given in Eq. (11).

Next an eigenanalysis is performed on the resulting covariance matrix C. Mathematically we are looking for eigenvectors x_j which satisfy

$$C\,x_j = \lambda_j x_j \tag{13}$$

for a scalar λ_j. The resulting eigenvectors are the basis spectra we seek. However, these eigenvectors must be sorted according to their corresponding eigenvalues before placement in the columns of the matrix A, as defined in Eq. (5). Sorting should be done such that the eigenvector that has the largest eigenvalue is placed in the first column of A, and the others are placed according to decreasing eigenvalues. This ordering ensures that the principal components which contribute most to the approximation are selected when we use less than the total number of basis spectra N_{hars}.

After the basis spectra are determined, the complete set of weights may be solved for by Gaussian elimination. Alternatively, least squares may be used to determine only those weights associated with the most important basis spectra, our usual case. Since the PCA-generated basis spectra are orthogonal to one another, the results of the two methods are identical.

5 MATCHING RESULTS

In the case of four or more basis spectra the GA-index and PCA approaches to determining basis spectra result in perceptually similar matches. In both cases, if the number of basis spectra equals the number of harmonics of the tone, an exact match can be made through ordinary additive synthesis of harmonic sine waves. In most matches tried to date, four or five basis spectra seem to be adequate for achieving excellent simulation. Even fewer basis spectra can be used if only reasonable approximations are desired. The GA-index approach generally fares better on these lower order matches.

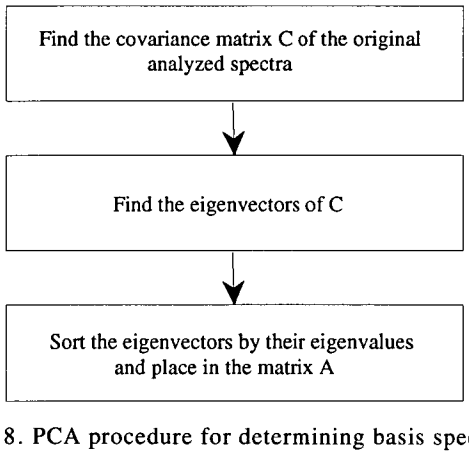


Fig. 8. PCA procedure for determining basis spectra.

5.1 GA-Index Matching Results

As mentioned earlier, we have found the GA to be most practical when it is used to choose basis spectra from the set of analysis spectra. The matches performed to date have been based on the sounds of a trumpet, a tenor voice, and a guitar. The trumpet was only remotely approximated when one basis spectrum was used. Of course, the use of more tables gave much better matches. Surprisingly a one-basis-spectrum match to the tenor voice did sound quite close to the original, while with two basis spectra, the match was almost perceptually indistinguishable from the original. The decay portion of the guitar tone was quite easy to capture, but its attack was more elusive, regardless of the number of tables. This was probably due to the guitar tone requiring a very large number (80 or so in this case) of upper harmonics to adequately represent its attack transient. Even so, matches using a relatively small number of wavetables clearly sounded "guitarlike." Overall, the results for the three instruments paralleled those found in our FM matching study, except for the one-basis-spectrum case for the tenor voice, where the GA-index result was notably superior to the corresponding one-carrier FM match.

Figs. 9–14 show amplitude-versus-time plots for the second and fourth harmonics of the trumpet, tenor, and guitar. Amplitude envelopes for the original tones are displayed along with one, three, and five basis spectra approximations to the original. With the trumpet, higher order matches were required to capture the shape of the original envelopes, especially on upper partials, such as the fourth harmonic.

The tenor's amplitude vibrato (tremolo) is a distinguishing characteristic, and it shows up very clearly in Fig. 12. The single-table match basically ignores this tremolo, while the three- and five-table matches model it quite well. Given the crudeness of the single-table match, it is rather remarkable that the resynthesis sounds as good as it does. This suggests that the tremolo of the tenor tone is only a by-product of the tenor's wide vibrato and is of secondary perceptual importance. The importance of frequency vibrato for voice sounds was also emphasized by Chowning in his FM model for soprano voice synthesis [27]. In a two-table GA match to the tenor, the tremolo was captured by periodically cross-fading between the two tables. This result is similar to that used in a previous vocal analysis and synthesis study [21].

The guitar is characterized by a very bright attack followed by a rather simple decay. The decay envelopes are modeled quite well by a single-table match, as shown in Figs. 13 and 14. However, as mentioned, the impulsive attack is much more difficult to mimic. Note

the spike which occurs near time zero in the original tone's fourth harmonic amplitude shown in Fig. 14. Only the five-table match manages a similar spike, and even its level is lower than the original.

Fig. 15 illustrates the average relative error [defined by Eq. (9)], plotted against the number of basis spectra used to match various tones. The graph shows that as the number of basis spectra approaches the number of

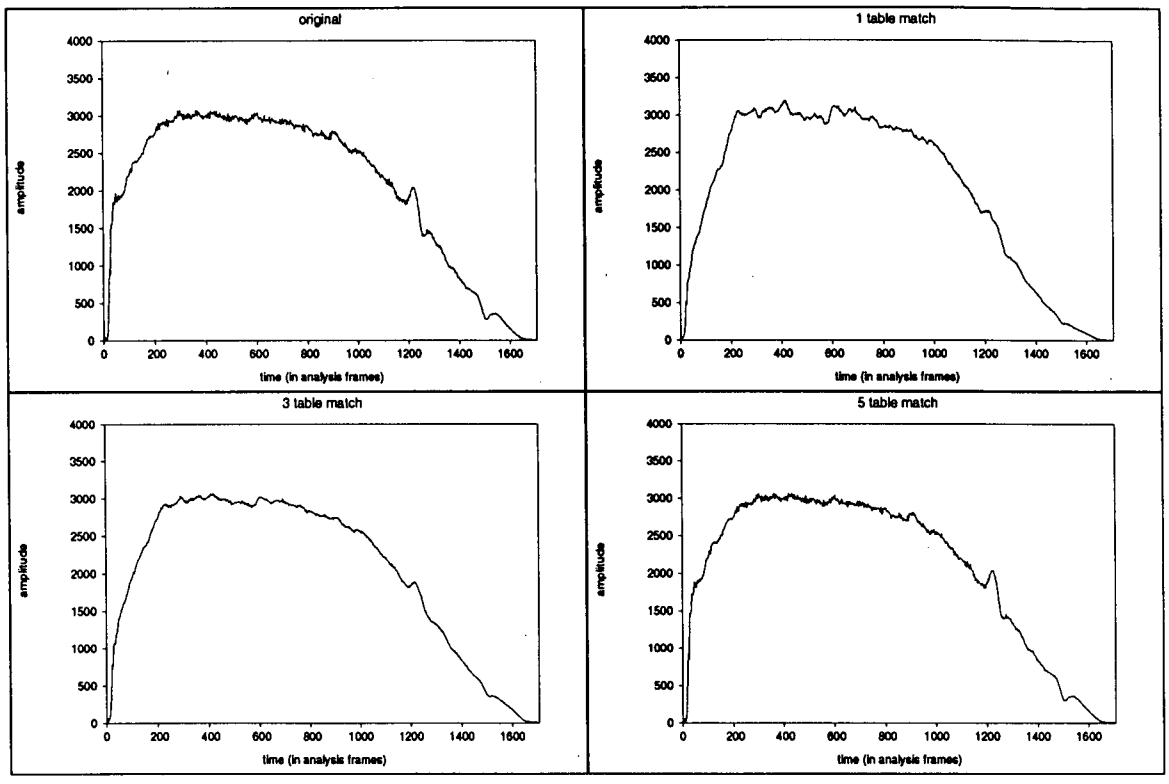


Fig. 9. Second-harmonic amplitude envelope of trumpet: Original and 1-, 3-, and 5-table GA-index matches. Duration is 2.4 s.

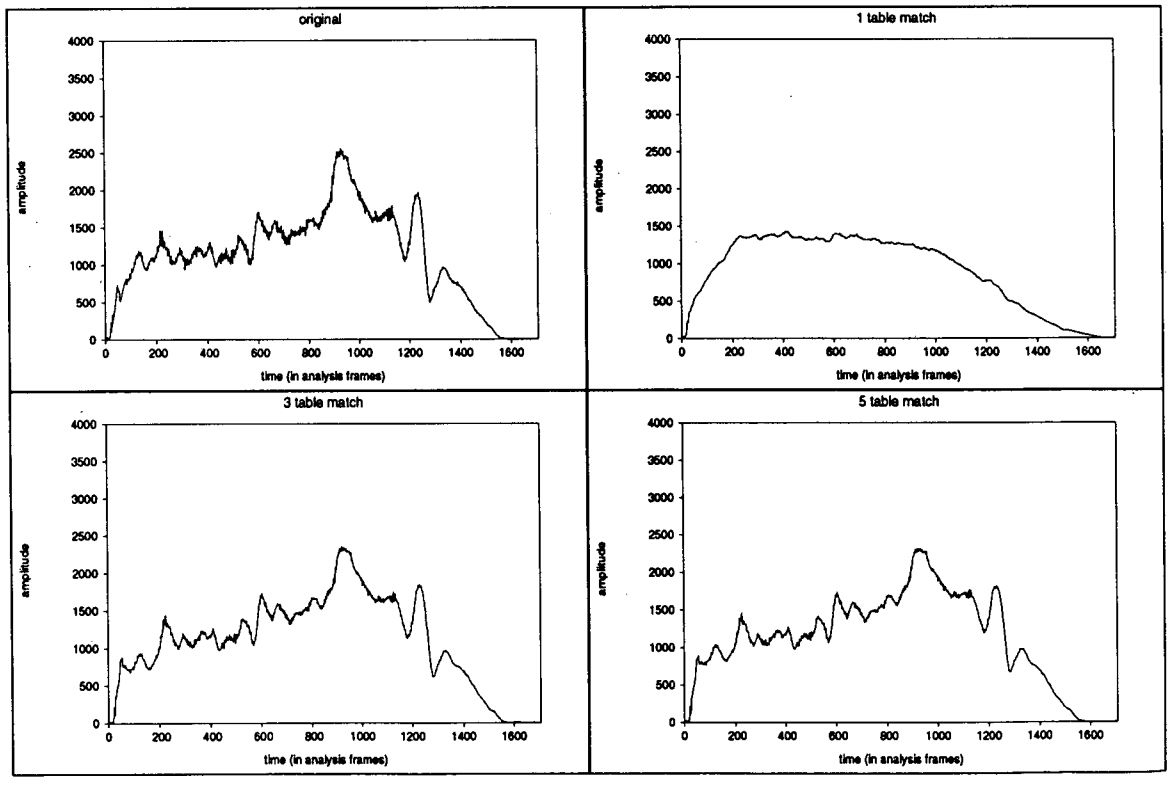


Fig. 10. Fourth-harmonic amplitude envelope of trumpet: Original and 1-, 3-, and 5-table GA-index matches. Duration is 2.4 s.

harmonics, the error does indeed tend to zero. These curves should not be used to compare the relative quality of matches for different original sounds, however. Our error measure should not be construed as an absolute measure of subjective quality, but it is usually indicative of how matches to a particular sound compare with one another. Even for the same sound, there is no guarantee that the result of minimizing the least-squares

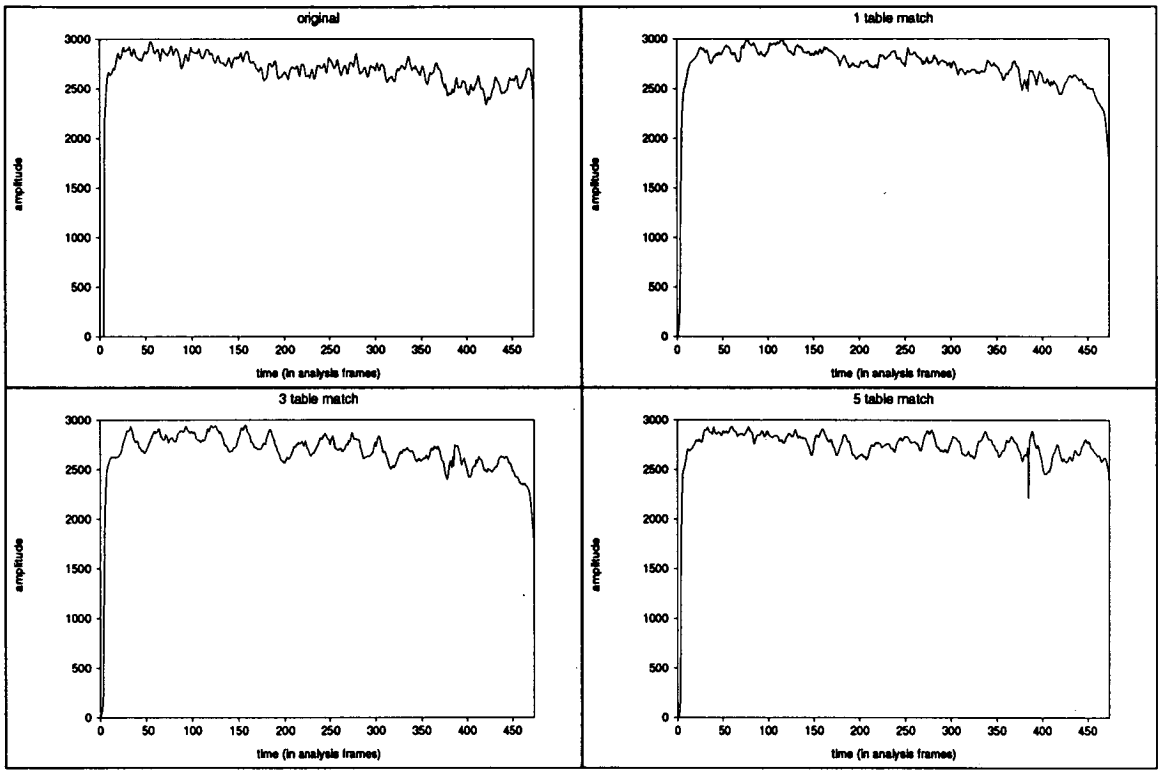


Fig. 11. Second-harmonic amplitude envelope of tenor voice: Original and 1-, 3-, and 5-table GA-index matches. Duration is 3.9 s.

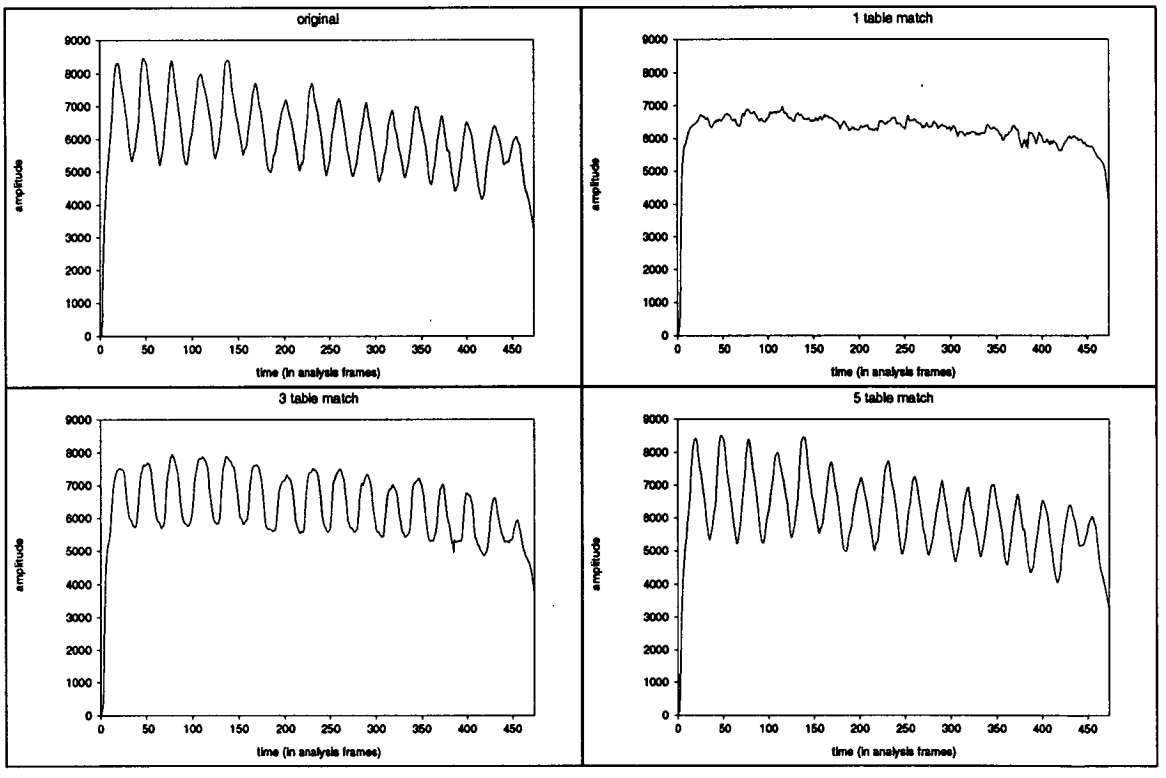


Fig. 12. Fourth-harmonic amplitude envelope of tenor voice: Original and 1-, 3-, and 5-table GA-index matches. Duration is 3.9 s.

error will give the best perceptual match.

For comparative purposes we include Fig. 16, which illustrates the average relative error that would occur using sine-wave additive synthesis with a restricted number of harmonics. The harmonics used are not necessarily the lowest, but rather are those which contribute the most to reducing the error. Comparison with Fig. 15 reveals that the matching error generally converges

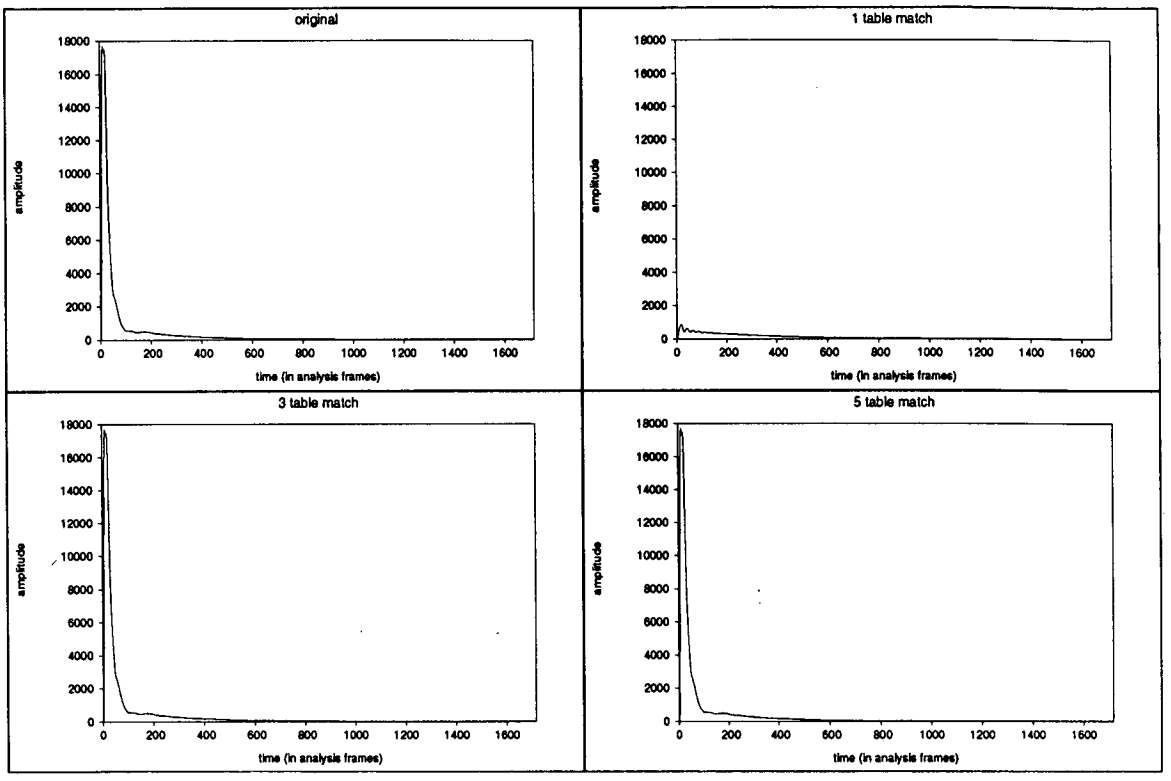


Fig. 13. Second harmonic amplitude envelope of guitar: Original and 1-, 3-, and 5-table GA-index matches. Duration in time is 8 s.

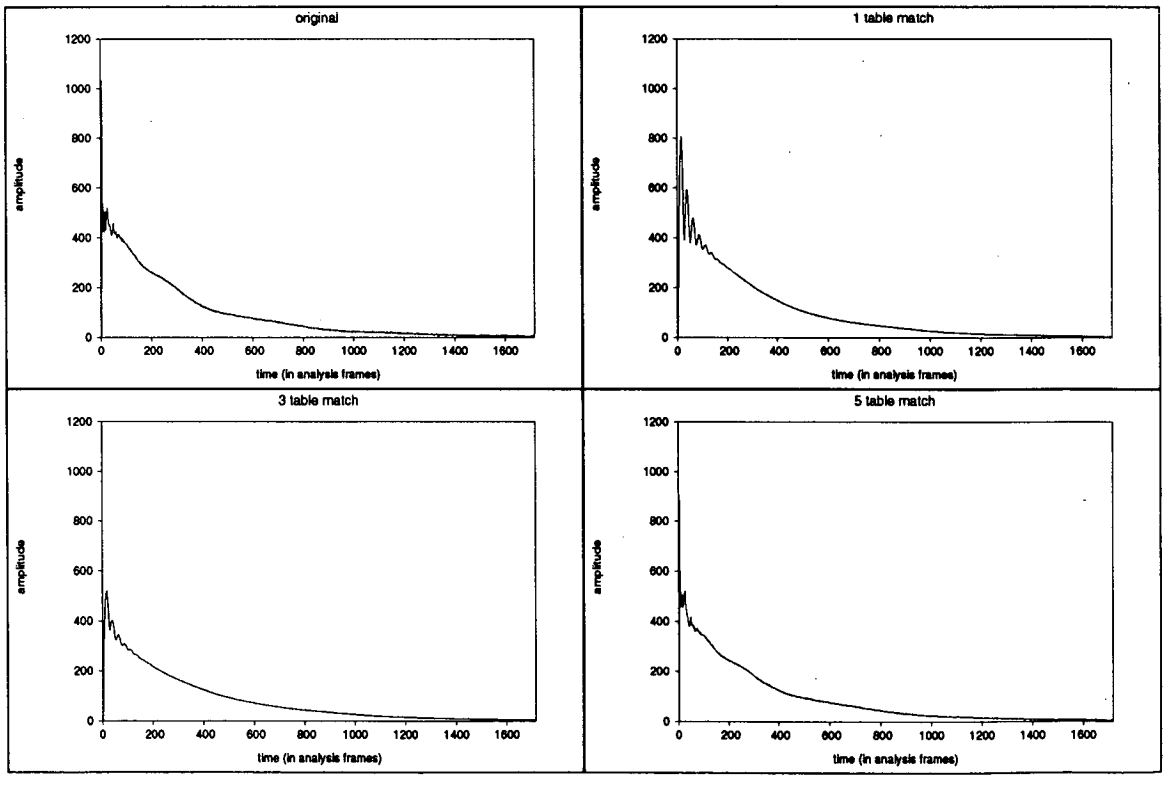


Fig. 14. Fourth-harmonic amplitude envelope of guitar: Original and 1-, 3-, 5-table GA-index matches. Duration is 8 s.

much more rapidly with wavetable matching synthesis than with sine-wave additive synthesis. The relatively fast convergence for the guitar tone shown in Fig. 16 is due to the fact that there are only a handful of harmonics of significant amplitude during its long decay.

The optimized parameters for a three-table match to the trumpet sound are given in Fig. 17. It shows plots of the basis spectra and amplitude envelopes $w_j(t)$ for this match. We can get a feel for what is happening in the match by examining these parameters. Note that during the initial attack of the tone, table 1, then table 3, and finally table 2 fade in successively. This corresponds to the brightening of the spectrum during the attack. Table 2 then dominates the tone's sustain portion. This table is in fact drawn from spectra in the middle portion of the tone. About halfway through the tone, the brightness begins to decrease, and we see a corresponding mixing of the tables. During this section, table 1's weight is negative, indicating that the lower harmonics are being partially canceled to offset the addition of both tables 2 and 3. Eventally table 3 emerges as the dominant basis spectrum. As the brightness of the tone wanes, table 3 cross-fades with table 1, the reverse of the opening trend.

The three tables were drawn from analysis frames 493, 1336, and 1575 (times 0.69, 1.88, and 2.22, respectively). At each of these time points, while the weight for the corresponding table is finite, the weights for the other tables are zero, since the least-squares procedure forced a perfect fit with the appropriate source table at each of these points. Fig. 18 plots error versus time for this three-table trumpet match. Note that the

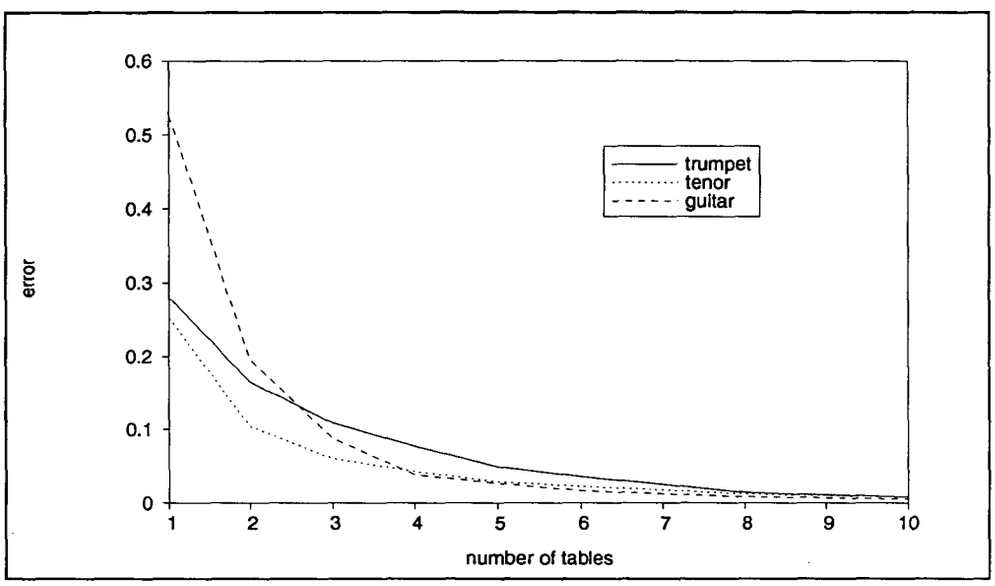


Fig. 15. Convergence of average relative error with increasing numbers of wavetables with the GA-index method.

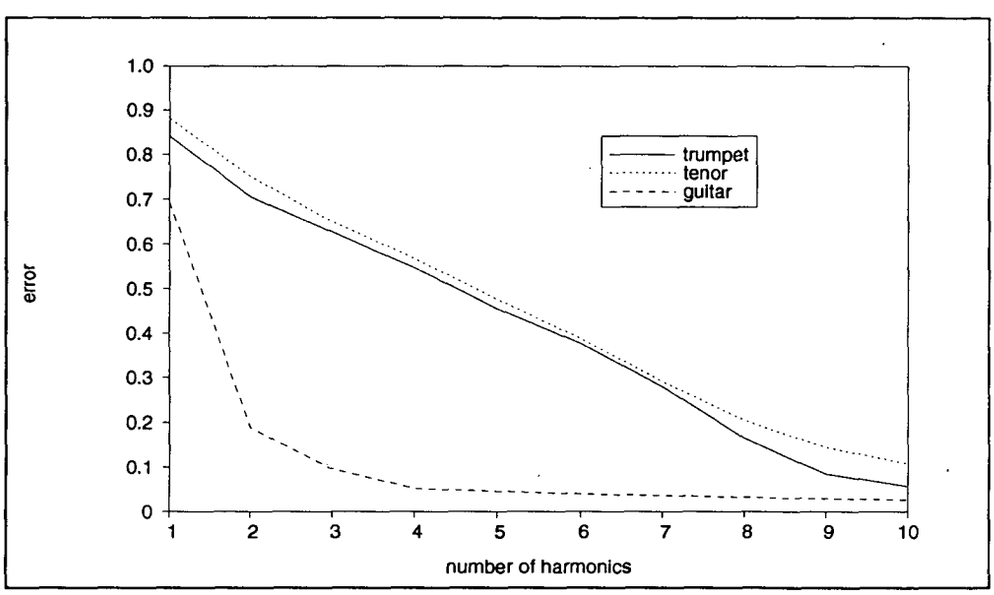


Fig. 16. Convergence of average relative error using sine-wave additive synthesis.

error goes to zero at the time points at which the basis spectra occur in the original sound, reflecting perfect matches at these points. Larger values of errors tend to occur during the transient attack and decay portions of the tone, where the spectral changes are more diverse. This example is suggestive of how basis spectra typically interact in their final mix.

5.2 PCA-Based Matching Results

Principal-components-based matching results are perceptually similar in character to those found in genetic matching, though they suffer from problems inherent in the underlying statistical approach. For our PCA trumpet simulations, this was primarily manifested as an excess of brightness in the release of the synthetic tone. Fig. 19 shows the relative error-versus-time curve for a PCA-based trumpet match where three wavetables were used. Fig. 19 should be compared to Fig. 18, which is for the corresponding GA match. Note that the relative error never goes to zero in the PCA match, since none of the basis spectra ever exactly matches a particular short-time spectrum of the original tone. Also note that the error is consistently low in the sustain portion of the tone, but is much higher during the attack and decay. Generally PCA decomposition of a sound will suffer from relatively large errors during the low-

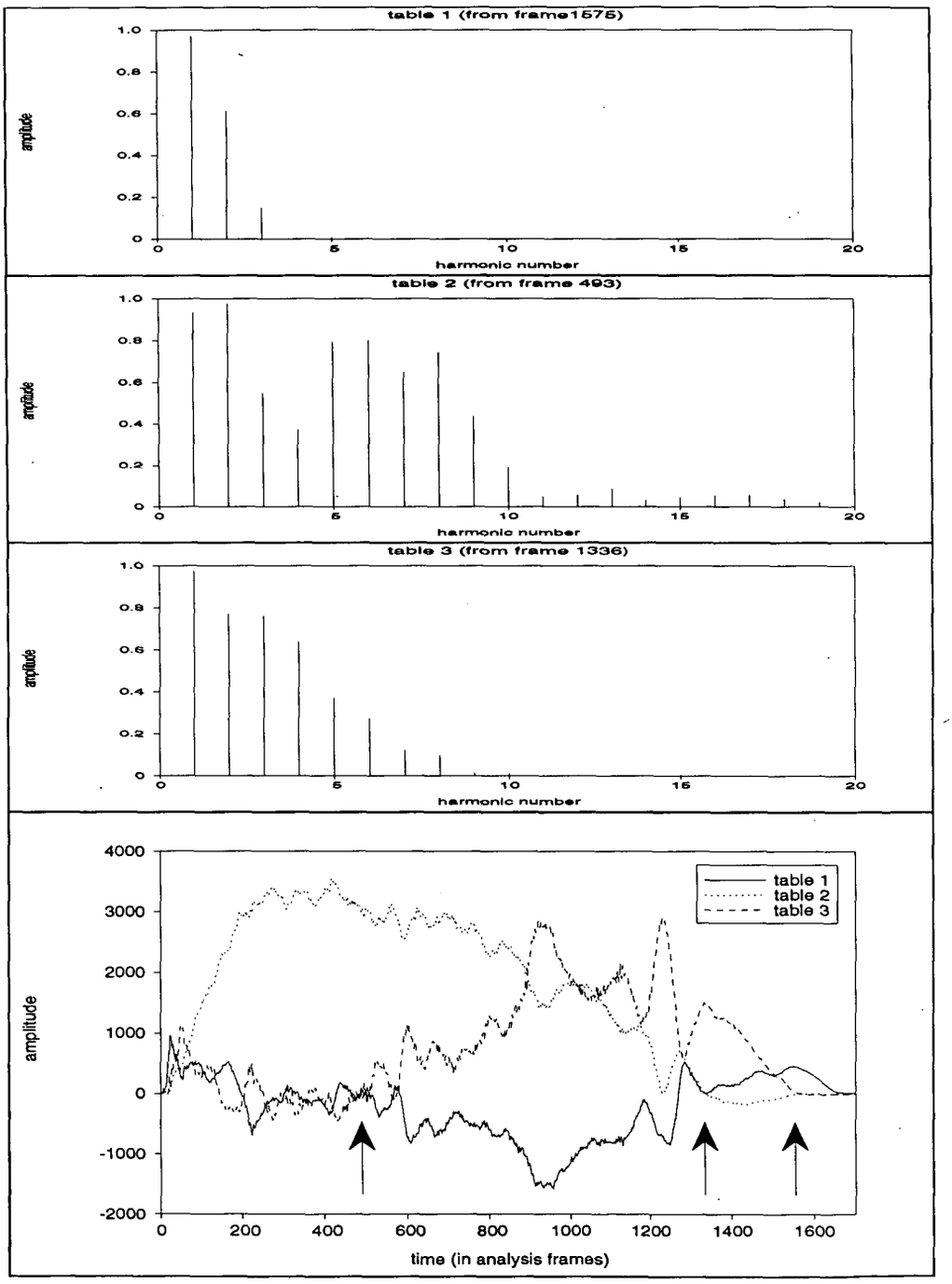


Fig. 17. Basis spectra and amplitude envelopes of a 3-table GA-index match for the trumpet. Arrows indicate times for which only one of the basis spectra is employed in the mix.

amplitude sections of the sound. This is due to the fact that most of the spectral variance will be caused by the highest amplitude portions of the tone. Thus the matching accuracy during lower amplitude sections will generally be sacrificed in order to match the higher amplitude sections better. While this problem might be obviated by using a logarithmic amplitude measure, we would then lose the linear additive synthesis feature which we require. Another possibility, which we have not checked out, would be to use more spectra from the low-amplitude portions of a sound than from the high-amplitude portions in our PCA analysis.

A statistical artifact also occurred in the case of the tenor voice. Fig. 20 illustrates the problem. We see that the first principal component with its time-varying weight tracks the tremolo of the tone. Examining the second basis spectrum and its amplitude envelope in isolation, we see that this component tracks the tones' rms amplitude, which is rather flat. Normally we might expect the first principal component to track this. However, for the tenor the variance of the tone's tremolo is several times greater than that of the tone's average amplitude. This fact leads PCA to decompose the tone with primary emphasis on its tremolo. Neveretheless, combining the first two principal components yields a shape very similar to that desired.

If we resynthesize the tenor tone with only the first principal component, the tone seems to turn on and off with each period of the tremolo. Though the tenor voice is clearly heard in the background of these modulated bursts, this is not what one hopes for in a good sounding match. However, if the second principal component is added, the match is suddenly very convincing, since the resulting sound now has both the correct tremolo and the correct overall spectral shape. This result is perceptually similar to that found by GA matching. In that case, however, the two basis spectra were cross-faded to emulate the periodically alternating spectrum of the tenor, rather than as an oscillation on top of a baseline spectrum.

Figs. 21–26 illustrate amplitude-versus-time plots for the second and fourth harmonics of the trumpet,

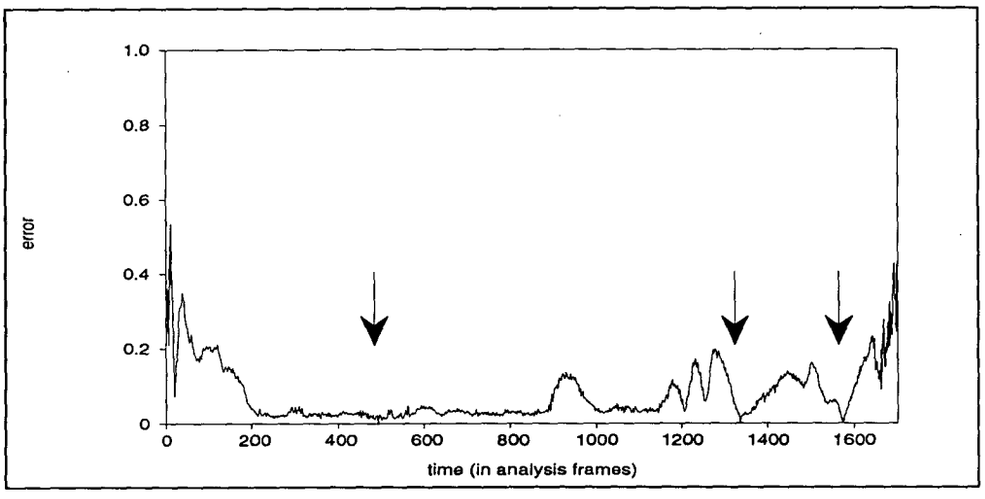


Fig. 18. Relative error versus time for a 3-table GA-index match for the trumpet. Arrows indicate times for which only one of the basis spectra is employed in the mix.

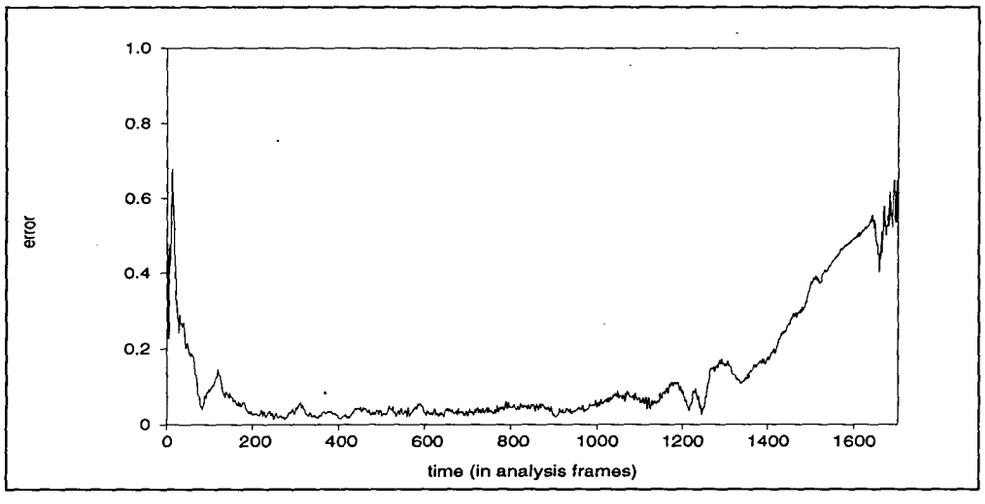


Fig. 19. Relative error versus time for 3-table trumpet match using PCA.

tenor, and guitar PCA matches. Amplitude envelopes for the original tones are displayed along with one, three, and five basis spectra approximations to the original. With the trumpet, these approximations are almost identical to those found by the GA-index method. The tenor's three- and five-table matches are also very similar to the GA result, while the single-table PCA match suffers from the isolated tremolo problem noted.

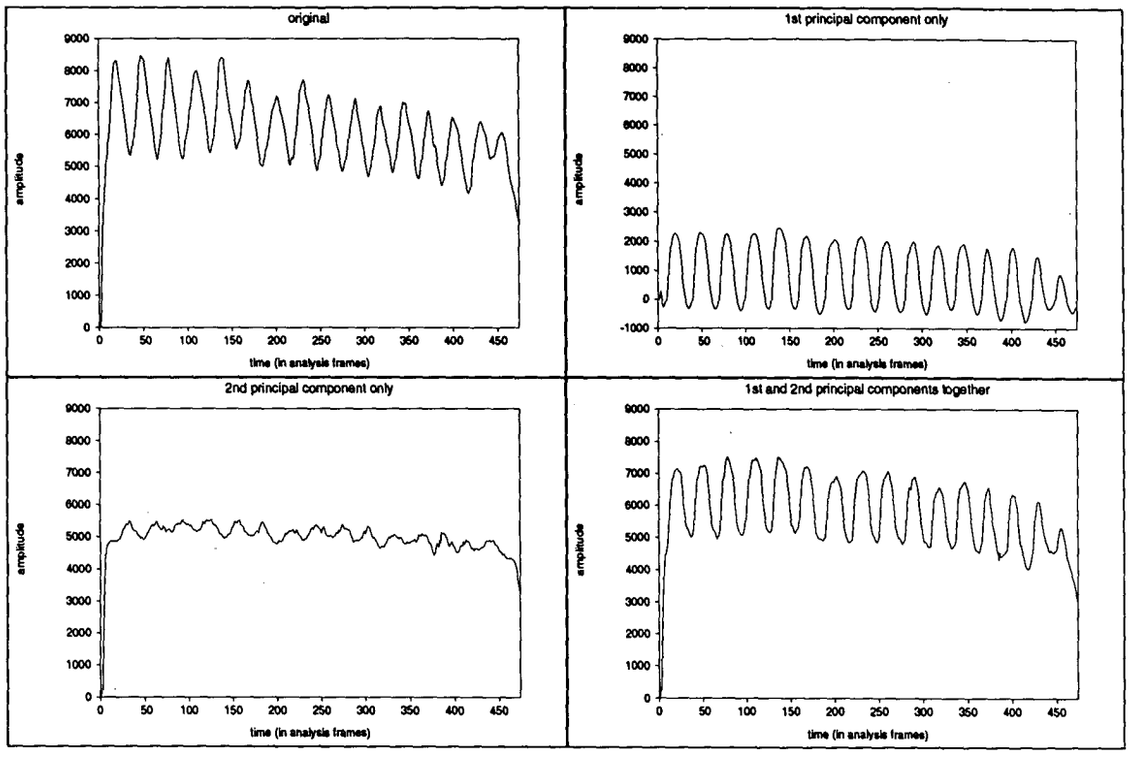


Fig. 20. Fourth-harmonic amplitude envelopes for the tenor voice: Original, first and second principal components individually, and first and second principal components combined.

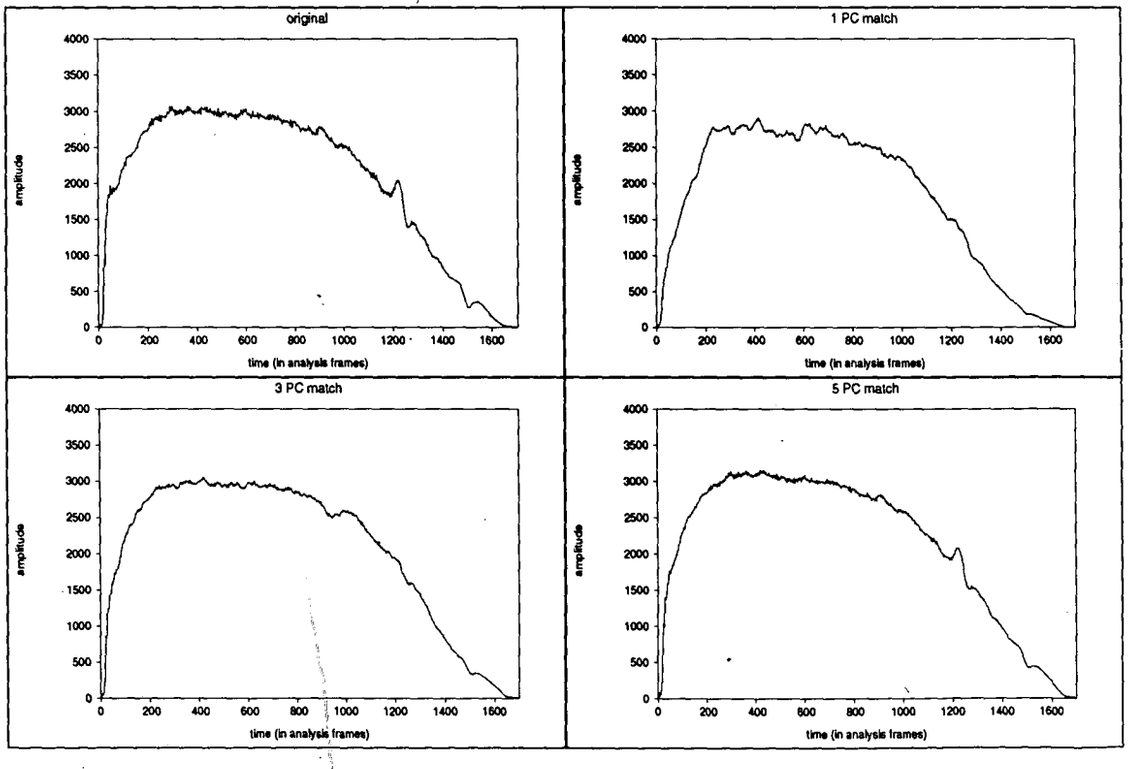


Fig. 21. Second-harmonic amplitude envelope of trumpet: Original and 1-, 3-, and 5-table PCA matches. Duration is 2.4 s.

Similarly to the GA result, the PCA guitar match has problems modeling the impulsive attack.

Results for a three-table PCA trumpet match are shown in Fig. 27. Note that the basis spectra contain negative harmonic components. Thus partial cancellation will take place even in the absence of negative weights. Curiously the weight envelope for table 1 is almost identical to the amplitude envelope of the second

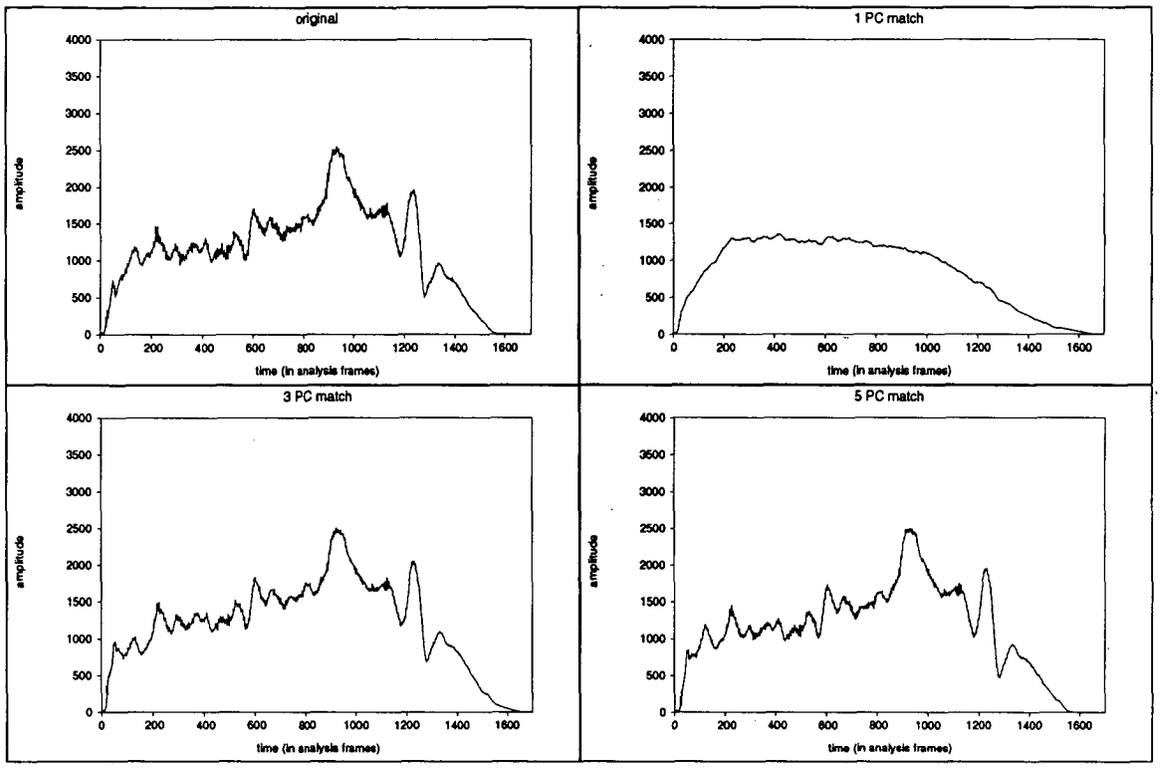


Fig. 22. Fourth-harmonic amplitude envelope of trumpet: Original and 1-, 3-, and 5-table PCA matches. Duration is 2.4 s.

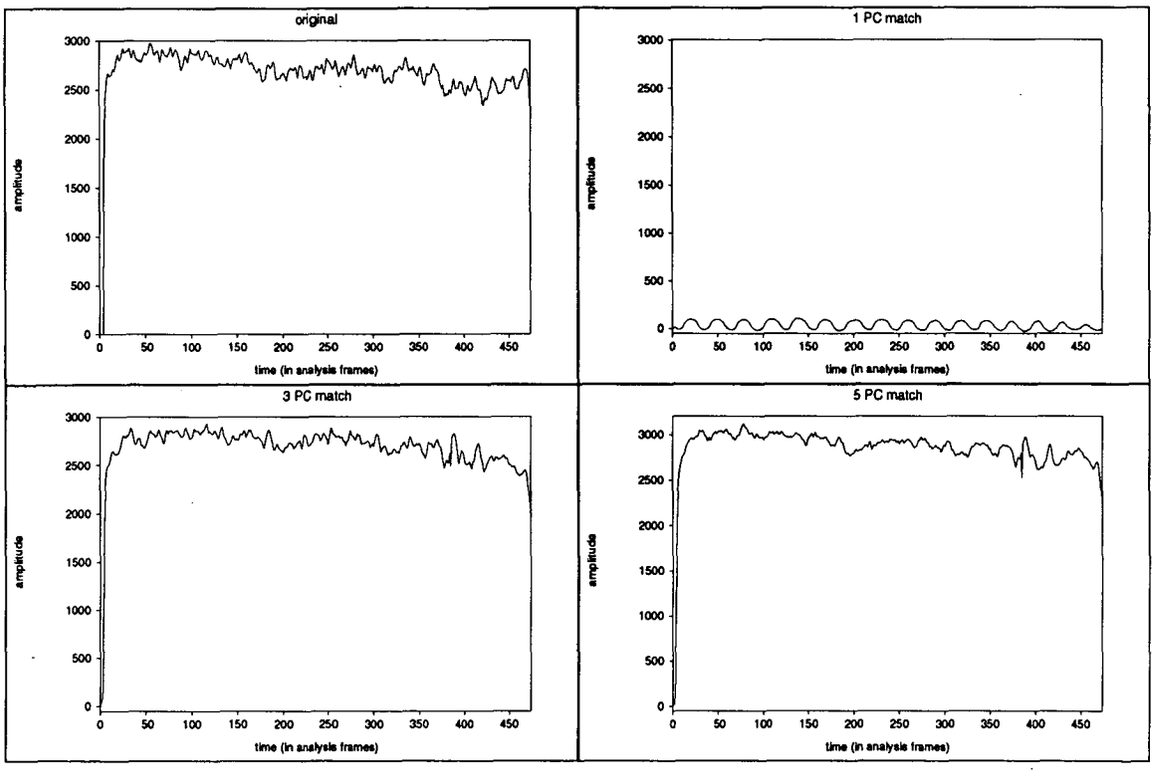


Fig. 23. Second harmonic amplitude envelope of tenor voice: Original and 1-, 3-, and 5-table PCA matches. Duration is 3.9 s.

harmonic shown in Fig. 21. However, the first table actually controls a broad spectrum of harmonics. The other tables, which, in general, have much smaller weights, sculpt the first basis spectra into the proper form. Aside from the negative components, the primary difference between the PCA basis spectra and the GA-selected basis spectra in Fig. 17 is the presence of upper harmonics in all the PCA tables. Even with can-

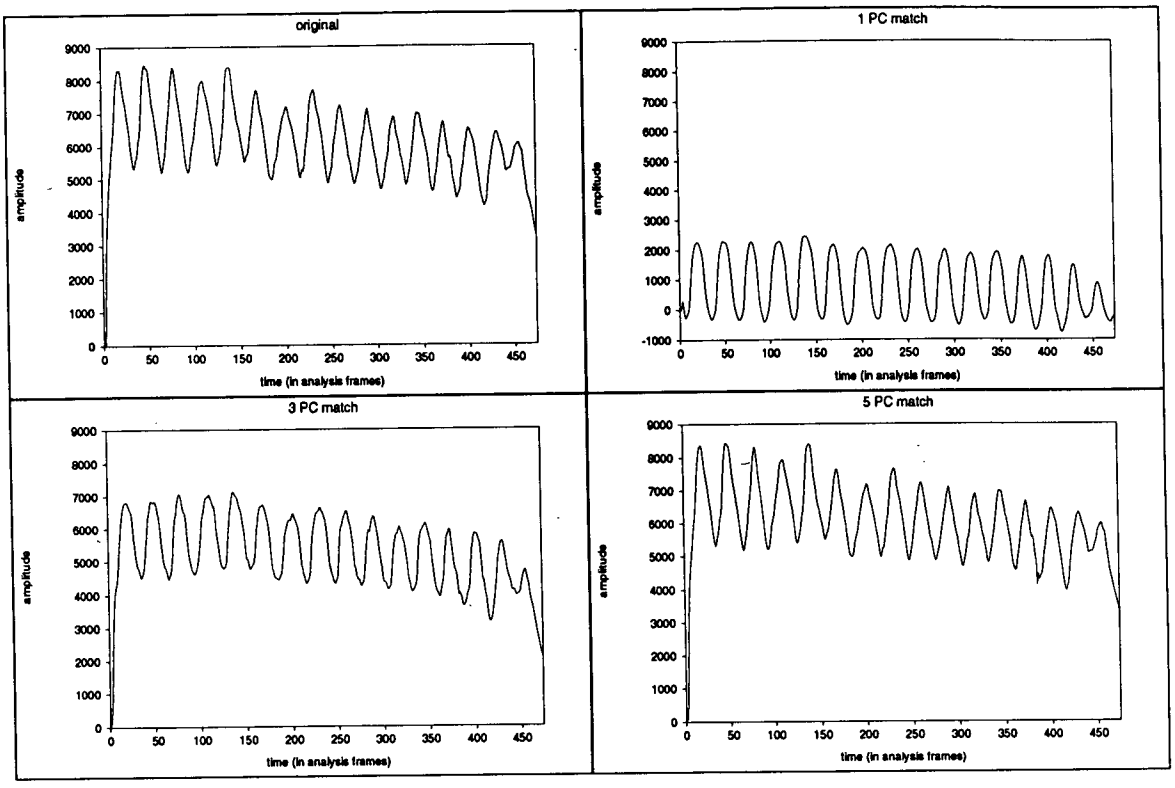


Fig. 24. Fourth-harmonic amplitude envelope of tenor voice: Original and 1-, 3-, and 5-table PCA matches. Duration is 3.9 s.

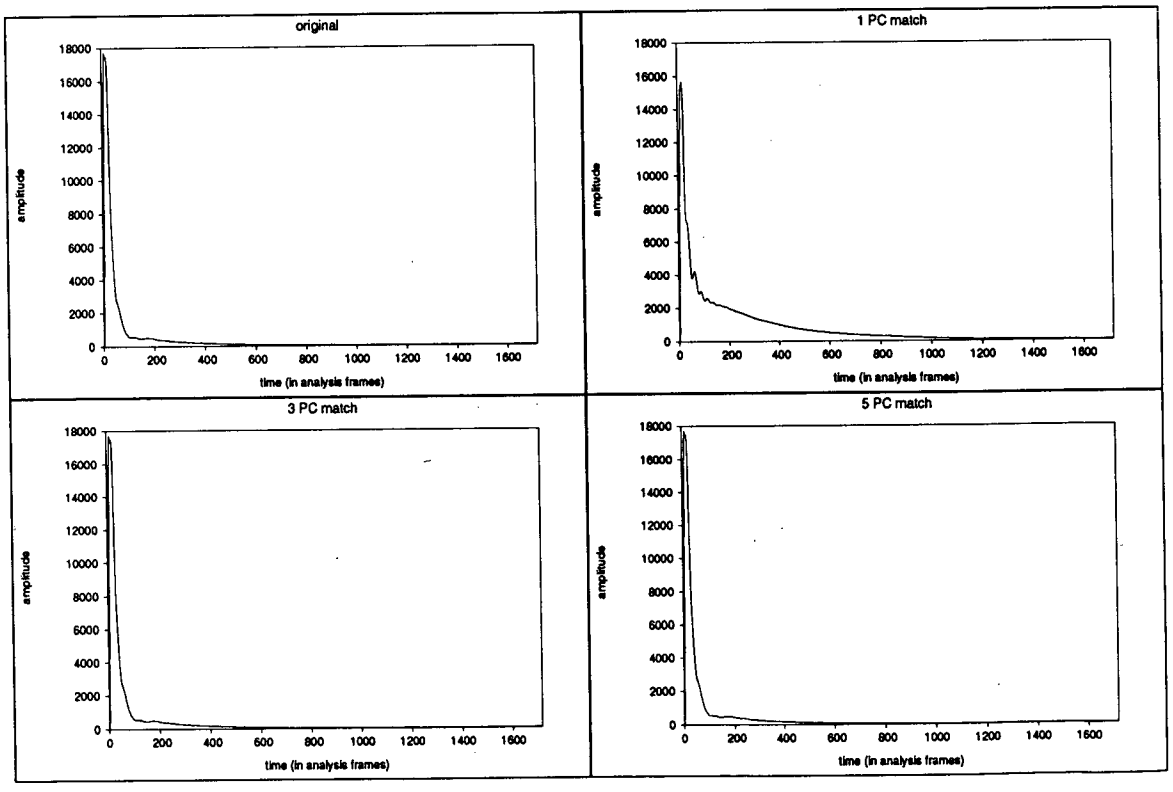


Fig. 25. Second-harmonic amplitude envelope of guitar: Original and 1-, 3-, and 5-table PCA matches. Duration is 8 s.

cellation, significant energy is bound to be left unchecked in the higher frequencies during the low-bandwidth spectral decay of the tone. This accounts for the excess brightness in the synthesized decay, due to the low-versus high-amplitude problem noted, which is characteristic of the PCA trumpet matches.

Fig. 28 shows the average relative error [defined in Eq. (9)] plotted against the number of principal component spectra used to match various tones. As in Fig. 15, when the number of basis spectra approaches the number of harmonics, the error tends to zero. However, the nature of the convergence is different. In general, the GA-index method leaves less error in its one- and two-table matches. With four or more tables, the errors are generally similar, although on average the GA method still wins. This suggests that for one or two wavetables the GA method is clearly superior. Moreover, the GA-index method shows more consistent improvement than the PCA method as the number of basis spectra is increased.

In conclusion, though statistically optimal, PCA should by no means be regarded as the best matching method. So far, results found by genetic selection of the analysis spectra are perceptually and numerically superior for relatively small numbers of wavetables. However, results, such as that found by PCA for the two-basis-spectra tenor match, may be useful in terms of decomposing the tone into components for further analysis or modification. Thus genetic and principal-components-based matching techniques offer different perspectives on the matching analysis of sounds.

6 CONCLUSIONS

We have explored two techniques for determining basis spectra and amplitude envelopes for resynthesizing tones via multiple fixed wavetable synthesis. Breaking down the matching processes into efficient, robust subprocedures was central to the success of both the GA-index and the PCA-based techniques. For four or more basis spectra, the GA-index and PCA methods gave similar results, but on average the GA-index results were markedly better. For less than four basis spectra, the GA-index approach was clearly superior. In the future we expect that these matching methods will be used to facilitate applications such as data reduction, data stretching, and synthesis by rule.

7 ACKNOWLEDGMENT

This material is based on work supported by the CERL Sound Group and the Computer Music Project at the University of Illinois at Urbana-Champaign. The work was facilitated by NeXT computers in the Computer Music Project at the School of Music of the UIUC and Symbolic Sound Corporation's Kyma workstation. The authors wish to thank the members of the CERL Sound Group, whose input and feedback have been invaluable in this work. These include Kurt Hebel, Carla Scaletti, Bill Walker, Kelly Fitz, and Richard Baraniuk. Thanks are also due to Lydia Ayers, Chris Gennaula, Camille Goudeseune, Chris Kriese, and Michael Hammond of the Computer Music Project for conversations related to this work.

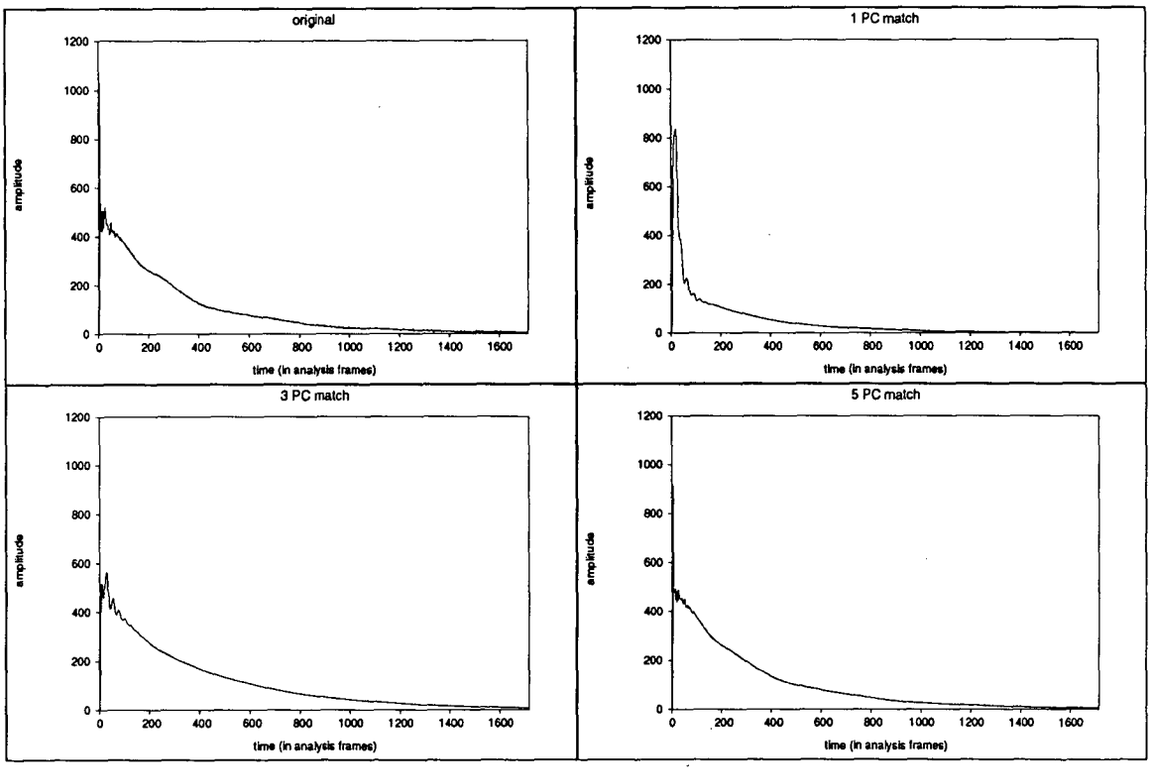


Fig. 26. Fourth-harmonic amplitude envelope of guitar: Original and 1-, 3-, and 5-table PCA matches. Duration is 3.9 s.

8 REFERENCES

[1] B. Atal and S. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," *J. Acoust. Soc. Am.*, vol. 50, pp. 637–655 (1971).

[2] C. Dodge, "*In Celebration*: The Composition and Its Realization in Synthetic Speech," in C. Roads, Ed., *Composers and the Computers* (A-R Editions, Inc., Madison, WI, 1985), pp. 47–74.

[3] J. W. Beauchamp, "Synthesis by Spectral Amplitude and 'Brightness' Matching of Analyzed Musical Instrument Tones," *J. Audio Eng. Soc.*, vol. 30, pp. 396–406 (1982 June).

[4] R. Payne, "A Microcomputer Based Analysis/Resynthesis Scheme for Processing Sampled Sounds Using FM," in *Proc. 1987 Int. Computer Music Conf.* (Int. Computer Music Assn., San Francisco, CA, 1987), pp. 282–289.

[5] N. Delprat, P. Guillemain, and R. Kronland-Martinet, "Parameter Estimation for Non-Linear Resynthesis Methods with the Help of a Time-Frequency Analysis of Natural Sounds," in *Proc. 1990 Int. Computer Music Conf.* (Int. Computer Music Assn., San Francisco, CA, 1990, pp. 88–90.

[6] A. Horner, J. Beauchamp, and L. Haken, "FM Matching Synthesis with Genetic Algorithms," *Com-*

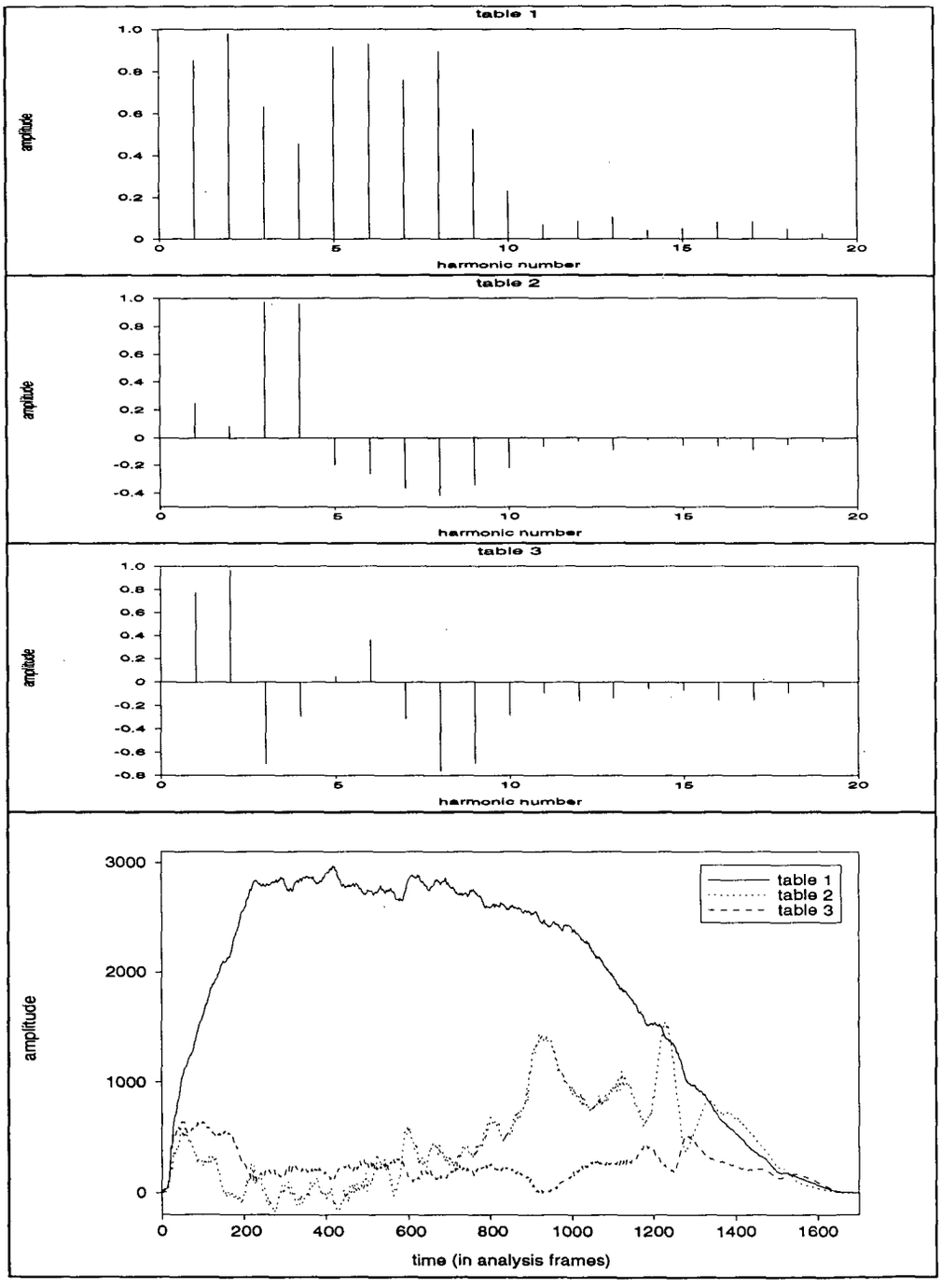


Fig. 27. Basis spectra and amplitude weight envelopes for a 3-table trumpet PCA match.

puter Music J., to be published, vol. 17 (1993).

[7] P. Kleczkowski, "Group Additive Synthesis," *Computer Music J.*, vol. 13, no. 1, pp. 12–20 (1989).

[8] J. Stapleton and S. Bass, "Synthesis of Musical Tones Based on the Karhuene-Loève Transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 305–319 (1988).

[9] M.-H. Serra, D. Rubine, and R. Dannenberg, "Analysis and Synthesis of Tones by Spectral Interpolation," *J. Audio Eng. Soc.*, vol. 38, pp. 111–128 (1990 Mar.).

[10] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA, 1989).

[11] J. Holland, *Adaptation in Natural and Artificial Systems*. (University of Michigan Press, Ann Arbor, 1975).

[12] G. Dunteman, *Principal Components Analysis* (Sage Publ., Newbury Park, CA, 1989).

[13] C. Chu, "A Genetic Algorithm Approach to the Configuration of Stack Filters," in *Proc. 3rd Intl. Conf. on Genetic Algorithms and Their Applications* (Morgan Kaufmann, San Mateo, CA, 1989), pp. 112–120.

[14] A. Horner and D. Goldberg, "Genetic Algorithms and Computer-Assisted Music Composition," in *Proc. 1991 Int. Computer Music Conf.* (Int. Computer Music Assn., San Francisco, CA, 1991), pp. 479–482.

[15] J. Stautner, "Analysis and Synthesis of Music Using the Auditory Transform," masters thesis, Dept. of Electrical and Computer Science, M.I.T., Cambridge, MA (1983).

[16] S. Zahorian and M. Rothenberg, "Principal-Components Analysis for Low Redundancy Encoding of Speech Spectra," *J. Acoust. Soc. Am.*, vol. 69, pp. 832–845 (1981).

[17] W. Hartmann, "Digital Waveform Generation by Fractional Addressing," *J. Acoust. Soc. Am.*, vol. 82, pp. 1883–1891 (1987).

[18] F. R. Moore, "Tablelookup Noise for Sinusoidal Digital Oscillators," *Computer Music J.*, vol. 1, no. 2, pp. 26–29 (1977).

[19] J. B. Allen, "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-25, pp. 235–238 (1977).

[20] R. McAulay and T. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 744–754 (1986).

[21] R. Maher and J. Beauchamp, "An Investigation of Vocal Vibrato for Synthesis," *Appl. Acoust.*, vol. 30, pp. 219–245 (1990).

[22] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, UK, 1989).

[23] G. Golub and C. Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, MD, 1983).

[24] M. Clark, Jr., D. Luce, R. Abrams, H. Schlossberg, and J. Rome, "Preliminary Experiments on the Aural Significance of Parts of Tones of Orchestral Instruments and on Choral Tones," *J. Audio Eng. Soc.*, vol. 11, pp. 45–54 (1963).

[25] K. Berger, "Some Factors in the Recognition of Timbre," *J. Acoust. Soc. Am.*, vol. 41, pp. 793–806 (1963).

[26] J. Grey and J. Moorer, "Perceptual Evaluations of Synthesized Musical Instrument Tones," *J. Acoust. Soc. Am.*, vol. 62, pp. 454–462 (1977).

[27] J. Chowning, "Computer Synthesis of the Singing Voice," in *Sound Generation in Wind, Strings, Computers* (Royal Swedish Academy of Music, Stockholm, Sweden, 1980).

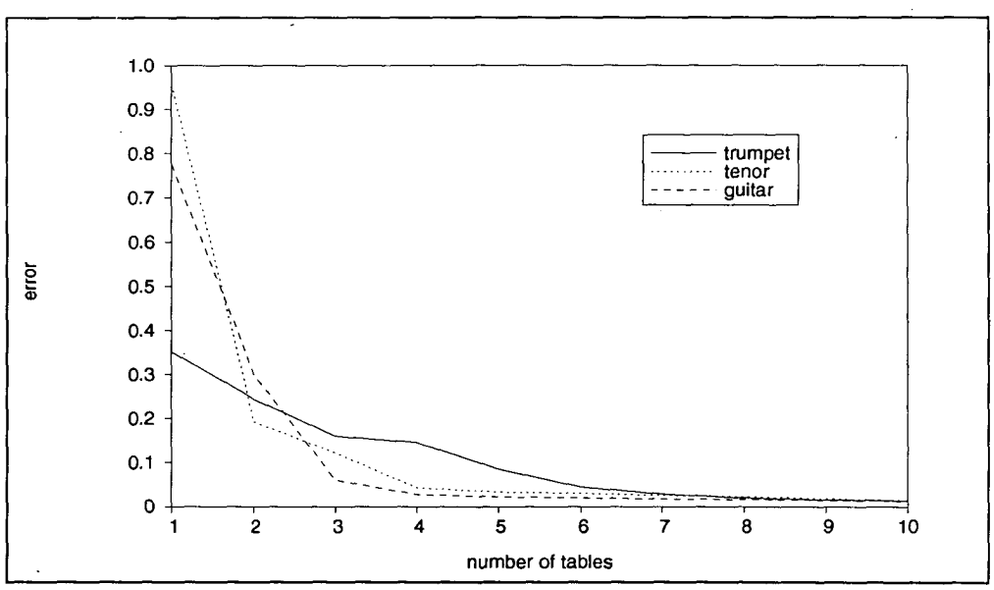
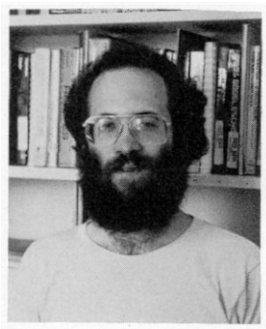
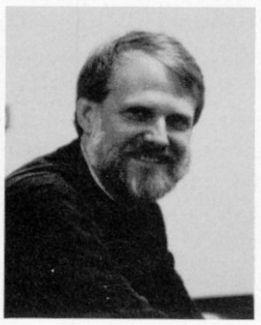


Fig. 28. Convergence of average relative error with increasing numbers of principal components.

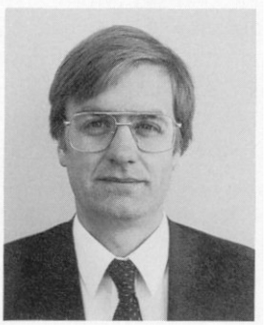
THE AUTHORS



A. Horner



J. Beauchamp



L. Haken

Andrew Horner was born in San Rafael, CA in 1964. He received a B.M. degree in music from Boston University and an M.S. in computer science from the University of Tennessee, Knoxville. He is currently finishing his Ph.D. studies in computer science at the University of Illinois at Champaign-Urbana.

At the University of Illinois, he is a researcher for the School of Music's Computer Music Project, the CERL Sound Group, and the Center for Complex Systems Research. He has also been supported by the Illinois Genetic Algorithm Laboratory. His primary research interests are in applying computational evolution to sound computation and computer-assisted composition.

•

James Beauchamp was born in Detroit in 1937. He received B.S. and M.S. degrees in electrical engineering from the University of Michigan during 1960–61 and a Ph.D. in electrical engineering from the University of Illinois at Urbana-Champaign in 1965. During 1962–65 he developed analog synthesizer equipment for the electronic music studio at UIUC under sponsorship of the Magnavox Company. He joined the UIUC electrical and computer engineering faculty in 1965 and began work on time-variant spectrum analysis/synthesis of musical sounds. During 1968–69 he was a research associate at Stanford University's Artificial Intelligence Project working on problems in speech recognition. Since 1969 he has held a joint faculty appointment in music and electrical and computer engineering at UIUC. In 1988 he was a visiting scholar at Stanford's Center for Computer Research in Music and Acoustics.

Dr. Beauchamp teachers courses at UIUC in musical acoustics, electronic music technology, audio, and computer music in both the School of Music and the Department of Electrical and Computer Engineering. Since 1984 he has directed areas of musical timbre characterization based on time-variant spectra of musical instrument tones, nonlinear/filter synthesis, and musical pitch detection.

He is a member of the Acoustical Society of America, a fellow of the Audio Engineering Society, and a member of the International Computer Music Association and its board of directors.

•

Lippold Haken was born in Munich, West Germany, and has lived mostly in central Illinois. He received a B.S. degree in 1982, an M.S. degree in 1984, and a Ph.D. in 1989 in Electrical and Computer Engineering from the University of Illinois. He is an Assistant Professor of Electrical and Computer Engineering at the University of Illinois, with research interests in audio signal processing, computer architecture, and user interface hardware and software.

He is leader of the CERL Sound Group, and has developed real-time audio signal processing hardware and software, focusing primarily on time-frequency analysis and synthesis of musical instruments. He is coauthor of a sophisticated music notation editor, Lime. He is also leader of the hardware design group for the Zephyr, a high-speed mainframe computer built at the University of Illinois Computer-based Education Research Laboratory. The Zephyr provides centralized program execution and data-keeping for thousands of simultaneous users on NovaNET, a real-time nationwide network used for computer-based instruction. Currently Dr. Haken is teaching a new project-oriented course that introduces the major areas of electrical and computer engineering to freshmen.