INTERNATIONAL SYMPOSIUM ON MUSICAL ACOUSTICS
AUGUST 28 – SEPTEMBER 1, 1992, TOKYO JAPAN 1B-1

# INVESTIGATION OF TIMBRE USING SPECTRAL MATCHING, MODELLING, AND INTERPOLATION

James Beauchamp, Andrew Horner, Chris Gennaula, and Camille Goudeseune
Computer Music Project, School of Music, University of Illinois at Urbana-Champaign
2136 Music Bldg., 1114 W. Nevada, Urbana, IL 61801 USA

Abstract: A number of methods for matching the outputs of wavetable synthesis algorithms to the time-variant spectra of musical instrument and voice sounds were investigated. Synthesis algorithms studied were common-modulator/multiple-carrier frequency modulation (FM), multiple fixed-wavetable synthesis, and nonlinear processor/filter (NLF) synthesis. Matching techniques used were a genetic algorithm (GA), least-squares solution, principal components analysis (PCA), and spectral centroid matching. For FM synthesis best results were obtained holding the indices fixed; a GA was used to select the best set of indices and carrier-to-modulator ratios. For fixed wavetable synthesis the PCA was used to compute orthogonal basis spectra for the tables. However, the GA method applied to selection of discrete spectra from a time-varying spectrum for fixed wavetables yields the best results, and 3 to 5 tables are usually sufficient for a good quality match. For certain instrument sounds with significant spectral centroid changes, the NLF technique with spectral centroid matching often yields an excellent match when 2 nonlinear modules are used, but the improvement with 3 modules is modest. Methods of modelling the spectral parameter behavior of vocal sounds and interpolation between different timbres are also discussed.

## 1. INTRODUCTION

Our overall objective is to develop synthesis models which allow us to control various aspects of sound quality over a large range of dynamic and pitch. Our point of departure is based on the set of conventional pitched musical instruments. One objective is to find models which can efficiently match the time-varying spectral characteristics of performed acoustic instrument sounds. Another is to determine maps of the variations of spectral parameters with time, pitch, and amplitude. Still another is to determine methods for interpolating between timbres which yield interesting intermediate results. It is hoped that this research will lead to an enhanced understanding of musical timbre in general.

## 2. SPECTRAL MATCHING RESEARCH

Spectrum matching of musical instrument tones is a fundamental problem in computer music. We have been exploring several models for synthesis where the parameters are optimized for synthesis which closely mimics the original sound. Generally, spectrum matching begins with a time-variant analysis of the original sound. Next, the synthesis parameters for "best fit" are determined with respect to the original sound. Finally, resynthesis of the sound is performed using the matched parameters. Thus far, the three synthesis models we have explored are a) frequency modulation (FM), b) additive synthesis with fixed wave tables, and c) nonlinear/filter synthesis.

As it turns out, each of these techniques is based on the linear sum of fixed or dynamic wave tables with time-varying weights. Techniques for determining the wavetable spectra include genetic algorithm (GA) optimization , principal components analysis (PCA) , and spectral centroid ("brightness") matching. Least-squares solution is utilized to determine the associated time-varying weights. The error computed by least-squares is also used to report on the goodness of fit and thus to guide the optimization process. What all of the methods have in common is that they utilize techniques to find the best wavetable spectra and their associated amplitude envelopes in order to match the original musical instrument time-varying spectra.

## 2. 1 Frequency Modulation Synthesis Matching

In attacking the FM problem, we decided to use a common modulator, multiple carrier model, since this model has been quite successful in the past. Since FM is a difficult nonlinear problem, we used the GA technique to attempt a solution. One thing we determined early on was that it was rather fruitless to allow the modulation indices to vary with time: Better results were obtained by keeping the indices fixed. The GA was then used primarily to determine the best set of fixed indices and integral carrier-to- modulator ratios for a given number of carriers. The synthesis technique could be implemented using multiple carrier FM, a very popular technique. However, as an alternative, it could be implemented using additive wavetable synthesis, where the table waveforms are those resulting from frequency modulation.

## 2.2 Fixed WaveTable Synthesis Matching

Another approach is use a model directly based on the linear sum of fixed spectra or wavetables. The object is to first find the best wavetable spectra for a given number of wavetables. Once these are found, the time-varying weights (envelopes) for the various tables can easily be found by the method of least-squares. Two methods which were successful for finding the wave table spectra were a GA spectrum indexing technique and a statistical technique based on PCA. Although one might contemplate using the GA to search the space of all possible spectra, this turns out to be a bad approach because the space is simply too sparse in terms of useful solutions. Instead, the search is confined to the set of discrete spectra contained in the sound to be matched. The GA, then, is focused on determining the indices (frame numbers) for the best spectra to serve as wavetable spectra. The PCA technique, on the other hand, is a straightforward matrix computation method which produces spectra best fitting the original sound in a statistical fashion. Since the PCA spectra are guaranteed to be an orthogonal set, an effective strategy is to take advantage of a closed-form matrix solution to arrive at the complete set of basis spectra and corresponding time-varying weights. Then, it is also guaranteed that as tables are added to the system, the error will continue to decrease until the number of tables equals the number of harmonics in the original spectrum, at which point the error will be zero.

One area for research is to determine an algorithm for predicting the perceptual difference between two similar sounds. So far we use a time average relative rms error criterion to calculate the goodness of fit (i.e., *fitness error*) between a synthetic time-varying spectrum and that measured from an original sound. However, a lower objective error does not always predict a smaller perceptual difference. In the future we hope to find a better objective fitness measure, both for guiding the matching process and for reporting the success of matching for a given synthesis model.

Figure 1 shows a comparison of average fitness error for the three methods discussed above, GA-FM, GA-index, and PCA, as functions of the number of tables used. The errors reported are averages for three different tones, those of a trumpet, a tenor voice, and a guitar. It is clear from these results that, in general, the GA-derived wave table approach is most effective. Contrary to what one might expect, PCA does not yield the best results, since in this case the basis spectra are derived to minimize average mean-squared error rather than average relative rms error (error normalized by value), which is our criterion for fitness. The FM results actually sound better than the numbers would indicate, but as one can see, this approach does not promote rapid convergence to the original time-varying spectrum. With each of these methods we find that three to five tables provide a useful result.

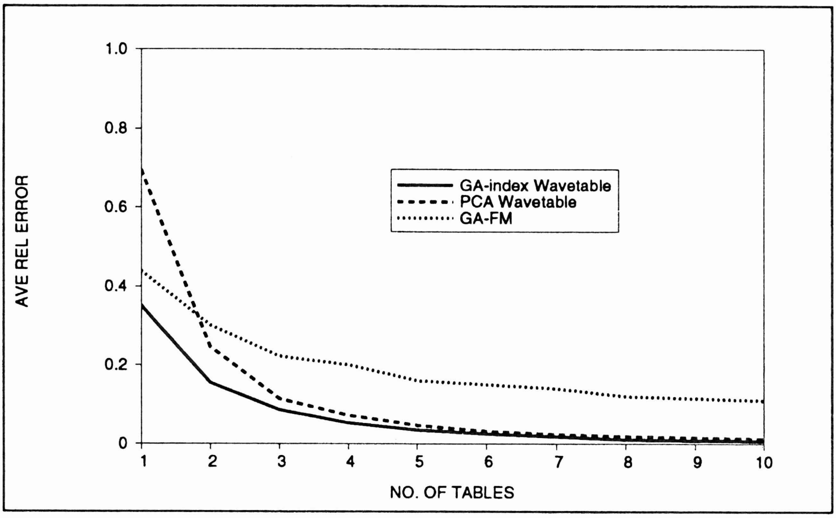


Figure 1. Comparison of matching error vs. number of tables for three different spectrum matching techniques

## 2.3 Nonlinear Processor/Filter Synthesis Matching

The last approach we tried is an extension of the nonlinear processing/filter (NLF) technique previously developed by one of us (Beauchamp). With this method we can attempt to match a dynamically-changing spectrum with a single NLF module. The dynamic spectrum is achieved by using the table as nonlinear processor of an amplitude-controlled sine wave. An additional high pass filter is used at the output to enhance brightness and improve the regularity of control. Whereas the contents of the table are determined by a "target spectrum" (generated by the table when the amplitude of the input sine wave equals unity), the instantaneous amplitude of the sine wave is calculated in order that the output spectrum matches the spectral centroid ("brightness") of the original sound. Then, a time-varying weight is applied to the table in order to force a least-squares match between the synthetic and original signals.

The extension of this technique consists of adding one or more modules to the basic method, each with its own target spectrum, sine wave amplitude control function, and post-multiplier function. The output of the modules is then summed and applied to the high pass filter. The process of generating additional modules is an iterative one. After the parameters of the first module are determined, the difference between the time-varying original spectrum and the predicted NLF spectrum are computed, and this residue is fed back into the NLF analysis process, which determines the parameters in the same manner as determined for the first module. Usually, we find that only one or two modules are needed to provide a perceptually close match to the original and that the addition of a third module does not improve the match sufficiently to justify the increased computation. Typical average relative rms errors are 0.2 for 1 module and 0.15 for 2 modules. It appears that the NLF technique has advantages in terms of the model's consistency over an extended range of pitches and dynamics. Also, the control functions, which are strongly related to brightness and loudness, should be advantageous for intuitive higher level control.

## 3. MODELLING SPECTRAL PARAMETER BEHAVIOR OF VOCAL SOUNDS

Several bass voice tones have been collected with pitches ranging from F2 to D4 at intervals of a minor third and recorded at six dynamic levels. Because of significant vibrato content the McAulay-Quatieri (MQ) technique was used to provide a time-varying spectrum analysis. The most revealing data display that we have discovered so far is an overlay graph of amplitude vs. frequency for each partial. This display method illuminates the formant characteristic of the voice. We hope to use this tool to find a generalized filter function for this voice as well as rules for varying the filter parameters as a function of pitch and dynamic.

## 4. INTERPOLATION BETWEEN DIFFERENT TIMBRES

One of us (Goudeseune) is investigating alternative ways to interpolate timbre in the domain of time-variant spectra. Given two or more spectra of equal duration, one may compute their weighted sum to produce an intermediate time-variant spectrum. However, differences arise as to how this sum should be defined. For example, harmonic envelopes can be defined parametrically, and weighted sums can be performed on the parameters so as to exploit corresponding features of envelopes from two different sounds. Moreover, linear measure preservation can be combined with techniques for two-dimensional linear interpolation. The thrust of this project is to explore how new "measure-preserving" methods of interpolation can be used to provide natural interpolations between radically different sounds.

## 5. SUMMARY

We expect that the results of these studies will have applications in performance synthesis in a variety of contexts. Precision matching allows us to estimate parameters needed for emulation of acoustical musical instrument sounds. We have tested three different techniques for matching sounds and have found the GA-index and extended NLF algorithms to the best methods. Modelling allows us to provide appropriate behavior across wide pitch and dynamic ranges, and here filter models appear to be most promising. The ability to crossfade among timbres is facilitated by enhanced interpolation techniques using measures which preserve temporal and spectral shapes across timbral boundaries.