SYNTHESIS OF TIMBRAL FAMILIES BY WARPED LINEAR PREDICTION†

by

Paul Lansky
Music Department
Princeton University
Princeton, New Jersey 08544

Kenneth Steiglitz
EECS Department
Princeton University
Princeton, New Jersey 08544

ABSTRACT

A central problem in the production of music by computer is how to generate sounds with similar but different timbres; that is, how to synthesize a "family" of instruments that are perceived as distinguishable but similar in timbre. This paper describes a method for synthesizing a family of string-like instruments that proceeds as follows: First, a few actual violin notes are analyzed using linear prediction. Second, a bilinear transformation is applied to the linear prediction model on synthesis, using a recently described algorithm. This yields a computationally efficient way to generate a virtual string orchestra, with violin-, viola-, violoncello-, and bass-like timbres. A realization of a 4.7-minute piece composed by P. Lansky will be played.

1. INTRODUCTION

We consider here the following problem faced by a composer using a digital computer for realization of his piece: He wants to generate sounds which are recognized as belonging to distinct families (such as strings, brass, etc.), and at the same time he wants different instruments in each family (such as violin, viola, etc.) to be clearly distingishable from each other. Since he obviously wants also to have more or less complete control over pitch, loudness, vibrato, and other features, the overall problem is quite complex.

This paper is devoted to describing an approach to this problem combining linear predictive coding and frequency warping that was implemented and used for the performance of a "string" piece written by one of the authors, P.Lansky. The method involves the following steps:

(a) A few typical notes (perhaps a tune) played on one member of the family are recorded and digitized.

(b) A linear predictive analysis is performed on the notes, and the results stored in large scale memory (on a disk, for example). Each frame represents an all-pole synthesis filter with transfer function 1/D(z).

(c) Synthesis is carried out using the general purpose program MUSIC 4BF[1], which in effect acts as an orchestra. Each instrument is synthesized by using the "frame filters" with transfer functions 1/D obtained in step (b) by linear predictive analysis, except that the excitation signals are determined by the instantaneous pitch requirements of the composer, and the filters used have undergone frequency transformations to change the effective timbre of the instrument.

In the rest of this paper we will fill in the details of these three steps, as we carried them out for the string piece.

2. Recording and Digitization

The starting material was a tune of 10 notes, lasting about 11 seconds, and played on a violin by Cyrus Stevens. Analog-to-digital conversion was done with a sampling rate of 14 kHz and 12 bits. This was the only sound used to produce the entire piece.

3. Linear Predictive Analysis

The covariance method [2] (with no window) was used for the linear predictive analysis, using 18 poles and 250-point (17.9 msec.) frames. The original signal was pre-emphasized by the highpass filter with transfer function $(1-z^{-1})$. The frames were not contiguous, but rather overlapped by 125 points, or 50%. That is, each new frame consisted of the last half of the previous frame followed by the first half of the next. This amount of overlap enables us to use linear interpolation of predictor coefficients without instability.

As is well-known, the covariance method can produce unstable frame filters. In this case, in fact, more than half the frames did have 1 or 2 poles outside the unit circle. One advantage of dealing with a such small amount of material is that it is not too expensive to factor every denominator D(z). This was done and any poles at radius R>0.998 were then moved to a radius of 0.998.

4. Excitation for Synthesis

A pitch contour was also obtained, using the algorithm in [3]. Deviations in pitch as much as ±1/3 of a semi-tone were observed. A semi-tone is

† This work was supported by NSF Grant ECS-7916292 and U.S. Army Research Office-Durham, Grant DAAG29-79-C-0024.

a frequency ratio of $2^{1/12}$, or about 6%, so the observed vibrato was as much as $\pm 2\%$ in pitch. Furthermore, the higher formants of the linear predictive analysis moved in a way correlated with the pitch, possibly because the higher formants were tracking harmonics of the excitation signal. This suggested that the pitch contour associated with a particular note be incorporated in the synthesis, so that the original correlation of pitch variation with formant motion be preserved. We accomplished that by transposing the original pitch contour to the desired pitch, and using the transposed pitch contour with the frame filters corresponding to the original pitch contour. Preserving the correlation between vibrato in pitch and formant structure seems to improve the realism of the synthesis.

The excitation signal itself was obtained by the method of [4], which uses a table lookup of a closed-form expression for a harmonic series with a flat spectrum. Its spectrum was then shaped by a 2-pole bandpass filter always centered at the fundamental, with a bandwidth of about 1500 Hz. The resultant harmonic series then descends in amplitude monotonically, and is a gross approximation to the spectrum of a triangular waveform, which one might guess is a good source for a bowed instrument.

In addition to transposing the original pitch tracking data, we also subjected the pitch to random deviations of ± 0.1 semi-tone. Seven such versions of each instrument, with different phases, were added to create the effect of an orchestral section, rather than a single instrument.

5. Warping of the Synthesis Filter

In [5] an efficient method is described for realizing the filter with transfer function

$$1/D'(z) = 1/D(A(z)) \qquad (1)$$

where $A(z)$ is the transfer function of any causal filter. To transform (warp) the frequency axis we choose the allpass function

$$A(z) = (d+z^{-1})/(1+dz^{-1}) \qquad (2)$$

If the original frame filter has transfer function

$$1/D(z) = 1/(1+\sum_{i=1}^{L} b_i z^{-i}) \qquad (3)$$

the transformed filter has the structure shown in Fig. 1, where the new coefficients are obtained by the simple calculation

$$\begin{aligned} b'_L &= b_L \\ b'_i &= b_i + db'_{i+1} \qquad i=L-1,\dots,1 \qquad (4) \\ c_L &= 1/(1+db'_1) \end{aligned}$$

and the function $B(z)$ is defined by

$$B(z) = A(z)-d \qquad (5)$$

$B(z)$ has no feedthrough term, so the signal flow graph in Fig. 1 has no delay-free loops and is

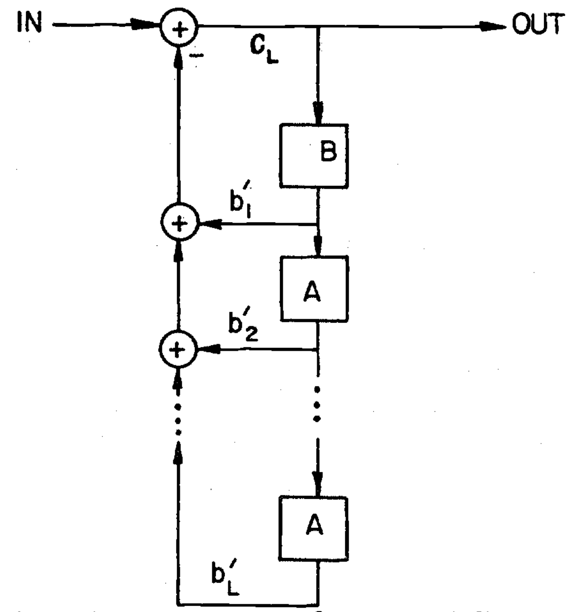

Figure 1 The structure of the warped filter.

therefore computable.

The effect of the tranformation is to map the frequency axis by

$$\phi(\omega) = \omega - 2\tan^{-1}(d \sin\omega/(1+d\cos\omega)) \qquad (6)$$

Thus, if the original transfer function has a formant at $\omega = \omega_0$, the new transfer function will have one when $\phi(\omega)=\omega_0$. This shows that when $0<d<1$, formants are shifted up; and when $-1<d<0$, formants are shifted down. To a second-order approximation in ω, $\phi(\omega)$ is linear at the origin, with slope

$$\phi'(0) = (1-d)/(1+d) \qquad (7)$$

so for small ω, a formant at ω_0 will be shifted to approximately $[(1+d)/(1-d)]\omega_0$.

We can view the frequency warping process as an approximate way of scaling the psychoacoustic size of the instrument. For example, in our experience in applying this technique to speech, the lowering of a woman's vocal formants gives the impression of a larger, but still feminine source, while raising the formants gives the distinct impression of a young girl.

For the family of string instruments, we used the fact that a viola is tuned 7 semitones lower than a violin, a violoncello 19 semitones, and a doublebass 27 semitones. Assuming the instrument sizes and formant structures are related in roughly this way, we used the values of d obtained from

$$(1+d)/(1-d) = 2^{-7/12}, 2^{-19/12}, 2^{-27/12} \qquad (8)$$

or

$$d=-0.19946, -0.49958, -0.65259 \qquad (9)$$

(If we know a pair of frequencies we wish to be related by the mapping, we can use Eq. (6) to find d precisely).

It was noted earlier that the high formants obtained from linear predictive analysis of the violin tend to track the pitch harmonics. On synthesis of high violin notes, therefore, it was found effective to shift the formants (via the relation (6)) by the same ratio as the pitch, so as to keep the high formants lined up with the pitch harmonics. This is not critical at lower pitches, since the pitch harmonics then are more closely spaced.

6. Conclusions

The final piece gives the intended impression of performance by an ensemble of strings. There are some clear differences, however, which may or may not be musically significant. The diminuendos, for example, sound artificial because no attempt was made to change the excitation or formants with loudness.

A linear predictive analysis of a few notes of a musical instrument contains within easy reach a great deal of musically useful information. A flexible analysis - synthesis, combining a careful linear prediction with a synthesis program like MUSIC 4BF, enables the composer to "play" real-sounding instruments with the computer. This paper describes a method that allows him to alter the apparent size of such instruments in a methodical way. Much work remains to be done, in applying the method to families of instruments other than strings, and in refining the technique.

REFERENCES

1. H.S. Howe, Jr., Electronic Music Synthesis, Norton, New York, 1975.

2. J.D. Markel and A.H. Gray, Jr., Linear Prediction of Speech, Springer-Verlag, New York, 1976.

3. K. Steiglitz, G. Winham, and J. Petzinger, "Pitch Extraction by Trigonometric Curve Fitting," IEEE Trans. ASSP, Vol. ASSP-23, No. 3, pp. 321-323, June 1975.

4. G. Winham and K. Steiglitz, "Input Generators for Digital Sound Synthesis," J. Acoust.Soc. Amer., Vol. 47, No. 2, (Part 2), pp. 665-666, February 1970.

5. K. Steiglitz, "A Note on Variable Recursive Digital Filters," IEEE Trans. ASSP, Vol. ASSP-28, No. 1, pp. 111-112, February 1980.