Paul Lansky
Music Department
Princeton University
Princeton, New Jersey 08540

Kenneth Steiglitz
Electrical Engineering and Computer
Science Department
Princeton University
Princeton, New Jersey 08540

Synthesis of Timbral Families by Warped Linear Prediction

Introduction

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

In this article we will describe an approach to this problem that combines linear predictive coding and frequency warping and that was implemented and used for the performance of a "string" piece, *Pine Ridge*, written by P. Lansky. The method involves the following steps:

1. A few typical notes (perhaps a tune) played on one member of the family are recorded and digitized.
2. A linear predictive analysis is performed on the notes and the results stored in large-scale memory (e.g., on a disk). Each frame represents an all-pole synthesis filter with transfer function $1/D(z)$.
3. Synthesis is carried out with the general-purpose program Music 4BF (Howe 1975), which in effect acts as an orchestra. Each instrument is synthesized by use of the *frame filters* with transfer functions $1/D$ obtained in step 2 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.

This paper was originally presented at the 1981 IEEE International Conference on Acoustics, Speech, and Signal Processing, 30 March to 1 April 1981, Atlanta, Georgia.

Computer Music Journal, Vol. 5, No. 3, Fall 1981,
0148-9267/81/000045-05 $05.00/0

Recording and Digitization

The starting material for *Pine Ridge* was a tune of 10 notes lasting about 11 sec 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.

Linear Predictive Analysis

The covariance method (Markel and Gray 1976) (with no window) was used for the linear predictive analysis. Eighteen poles and 250-point (17.9 msec) frames were used. The original signal was pre-emphasized by the highpass filter with transfer function $(1 - z^{-1})$. The frames were not contiguous, but 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 enabled 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, more than half the frames did have one or two poles outside the unit circle. One advantage of dealing with such a 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.

Excitation for Synthesis

A pitch contour was also obtained by use of the algorithm in "Pitch Extraction by Trigonometric Curve Fitting" (Steiglitz, Winham, and Petzinger 1975). Deviations in pitch as much as $\pm 1/3$ of a semitone were observed. A semitone is 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 would be preserved. We accomplished that by transposing the original pitch contour to the desired pitch and by using the transposed pitch contour with the frame filters that corresponded to the original pitch contour. Preserving the correlation between vibrato in pitch and formant structure seemed to improve the realism of the synthesis.

The excitation signal itself was obtained using the method described by Winham and Steiglitz (1970), in which a table lookup of a closed-form expression for a harmonic series was used 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 descends in amplitude monotonically and is a gross approximation of the spectrum of a triangular waveform, which is probably 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 semitone. Seven such versions of each instrument, with different phases, were added to create the effect of an orchestral section rather than a single instrument.

Warping of the Synthesis Filter

In the Appendix to this article an efficient method is described for realizing the filter with transfer function

$$\frac{1}{D'(z)} = \frac{1}{D(A(z))}, \tag{1}$$

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

$$A(z) = \frac{d + z^{-1}}{1 + dz^{-1}}. \tag{2}$$

If the original frame filter has the transfer function

$$\frac{1}{D(z)} = \frac{1}{1 + \sum_{i=1}^{L} b_i z^{-i}}, \tag{3}$$

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

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

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

$$B(z) = A(z) - d. \tag{5}$$

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

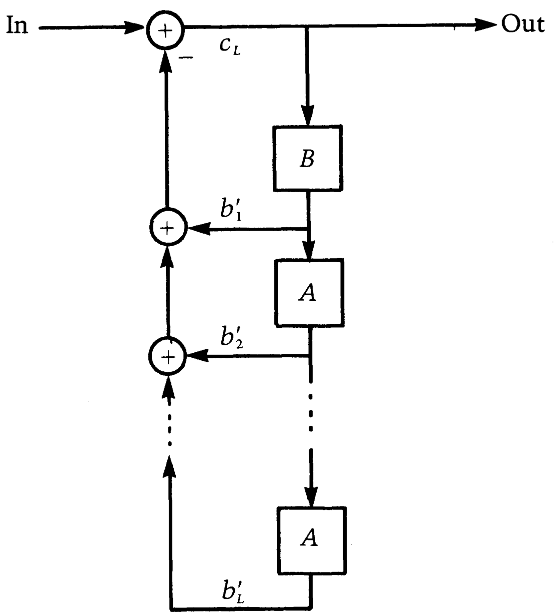
The effect of the transformation is to map the frequency axis by

$$\phi(\omega) = \omega - 2\tan^{-1}\left(\frac{d\sin(\omega)}{1 + d\cos(\omega)}\right). \tag{6}$$

Therefore, 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

$$\phi'(0) = \frac{1-d}{1+d}, \tag{7}$$

Fig. 1. The structure of the warped filter.



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 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's voice.

In the family of string instruments, a viola is tuned 7 semitones lower than a violin, a violoncello 19 semitones lower, and a double bass 27 semitones lower. We assumed that instrument sizes and formant structures are related in roughly this way and used the values of d obtained from

$$\frac{1+d}{1-d} = 2^{-7/12}, 2^{-19/12}, 2^{-27/12}, \tag{8}$$

$$d = -0.19946, -0.49958, -0.65259. \tag{9}$$

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

As noted earlier, high formants obtained from linear predictive analysis of the violin tend to track the pitch harmonics. On synthesis of high violin notes, therefore, we shifted 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 are more closely spaced.

Conclusion

Pine Ridge 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 diminuendi, 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 puts 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. In this article we have described a method that allows alteration of 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.

Appendix

A Note on Variable Recursive Digital Filters, by Kenneth Steiglitz[1]

Schüssler and Winkelnkemper [1970] note than when z is replaced by the low-pass-to-low-pass bilinear frequency transformation in the transfer function of a recursive digital filter, the resulting direct form structure has delay-free loops and is therefore not realizable without modification. John-

1. Reprinted from *IEEE Proceedings on Acoustics, Speech, and Signal Processing* 28: 111–112, 1980. Copyright© 1980 by the Institute of Electrical and Electronics Engineers, Inc. Reprinted with permission.

son [1976, 1979] gives two methods for computing the new coefficients in a realization which is in direct form, except for a factor of the form $(1 + dz^{-1})^k$, where k is the difference between the degrees of the original denominator and numerator. In [Johnson 1976] he shows that the new coefficients of both the denominator and numerator can be produced at the taps of a network similar to the frequency-warping network in [Oppenheim et al. 1971; Oppenheim and Johnson 1972]. In [Johnson 1979] an FIR network is used to recompute the coefficients. (Mullis and Roberts [1976] discuss the recomputation of coefficients in a state variable realization.)

In this correspondence we describe another method for realizing the transformed transfer function, one which preserves the all-pass substructure inherent in the bilinear transformation, and which results in a very fast coefficient recomputation.

We consider the transformation

$$z^{-1} \rightarrow \frac{d + z^{-1}}{1 + dz^{-1}} = A(z)$$

and define

$$B(z) = A(z) - d = \frac{(1 - d^2)z^{-1}}{1 + dz^{-1}} .$$

Given the transfer function

$$\frac{N(z)}{D(z)} = \frac{\sum_{i=0}^{M} a_i z^{-i}}{1 + \sum_{i=1}^{L} b_i z^{-i}} ,$$

the transformed transfer function is

$$\frac{\sum_{i=0}^{M} a_i A(z)^i}{1 + \sum_{i=1}^{L} b_i A(z)^i} .$$

The numerator can be implemented as it stands, as described in [Schüssler and Winkelnkemper 1970]. We therefore concentrate on the denominator: the difficulty is caused by the constant term in the sum. The following algebraic manipulation decreases by 1 the degree of the offending sum:

$$\frac{1}{1 + \sum_{i=1}^{L} b_i A^i} = \frac{1}{1 + (B + d) \sum_{i=1}^{L} b_i A^{i-1}}$$

$$= \frac{c_1}{1 + c_1 d \sum_{i=2}^{L} b_i A^{i-1} + c_1 B \sum_{i=1}^{L} b_i A^{i-1}}$$

where

$$c_1 = 1/(1 + db_1) .$$

If this is repeated L times, we get

$$\frac{1}{1 + \sum_{i=1}^{L} b_i A^i} = \frac{c_L}{1 + c_L B \sum_{i=1}^{L} b_i' A^{i-1}} .$$

Thus, we obtain the identity

$$1 + \sum_{i=1}^{L} b_i A^i = \frac{1}{c_L} + B \sum_{i=1}^{L} b_i' A^{i-1} .$$

Substituting $B = A - d$ and equating coefficients of like powers of A, we get the recursion relations

$$b_L' = b_L$$

$$b_i' = b_i + db_{i+1}' \qquad i = L - 1,\ldots,1$$

and

$$c_L = \frac{1}{1 + db_1'} .$$

The transfer function B has no feedthrough term, so this form of the denominator is directly realizable [Fig. 1].

When the numerator and denominator of the original filter are of the same degree, Johnson's form has the advantage of having the same filtering complexity as the original filter, although both numerator and denominator coefficients must be recomputed, whereas the present method requires recomputation of only the denominator coefficients. In the case of an all-pole transfer function, such as might arise in linear predictive coding, both methods result in a structure requiring more arithmetic to implement than the original.

A real advantage of the present method is the fast coefficient recomputation: the new b's can be found with only L multiplications, L additions, and one division, in contrast with the techniques described by Johnson, which appear to require $O(L^2)$ steps. The method also works without change for any causal A; B is defined by subtracting the constant term from A.

Acknowledgment

This work was supported by NSF Grants ECS-7916292 and GK-42048, and U.S. Army Research Office-Durham Grants DAAG29-79-C-0024 and DAHCO4-75-G0192.

References

Howe, Hubert S., Jr. 1975. *Electronic Music Synthesis.* New York: Norton.

Johnson, D. H. 1976. "Application of Digital-Frequency Warping to Recursive Variable-Cutoff Digital Filters." In *IEEE Electronics and Aerospace Systems Convention. Proceedings.* New York: IEEE, pp. 154A–154E.

Johnson, D. H. 1979. "Variable Digital Filters Having a Recursive Structure." *IEEE Proceedings on Acoustics, Speech, and Signal Processing* 27:98–99.

Markel, J. D., and A. H. Gray, Jr. 1976. *Linear Prediction of Speech.* New York: Springer.

Mullis, C. T., and R. A. Roberts. 1976. "Roundoff Noise in Digital Filters—Frequency Transformations and Invariants." *IEEE Proceedings on Acoustics, Speech, and Signal Processing* 24:538–550.

Oppenheim, A. V., D. H. Johnson, and K. Steiglitz. 1971. "Computation of Spectra with Unequal Resolution Using the Fast Fourier Transform." *Proceedings of the IEEE* 59:299–301.

Oppenheim, A. V., and D. H. Johnson. 1972. "Discrete Representation of Signals." *Proceedings of the IEEE* 60:681–691.

Schüssler, W., and W. Winkelnkemper. 1970. "Variable Digital Filters." *Archiv für Elektronik und Uebertragungstechnik* 24:524–525. (Reprinted in Rabiner, L. R., and C. M. Rader, eds. 1972. *Digital Signal Processing.* New York: IEEE Press.)

Steiglitz, K., G. Winham, and J. Petzinger. 1975. "Pitch Extraction by Trigonometric Curve Fitting." *IEEE Proceedings on Acoustics, Speech, and Signal Processing* 23:321–323.

Winham, G., and K. Steiglitz. 1970. "Input Generators for Digital Sound Synthesis." *Journal of the Acoustical Society of America* 47 (2, Part 2):665–666.