INTERPOLATION OF LPC SPECTRA VIA POLE SHIFTING

Vladimir Goncharoff and Maureen Kaine-Krolak

University of Illinois at Chicago, Department of Electrical Engineering and Computer Science
Mail Code 154, 851 S. Morgan St., Chicago, IL 60607-7053 (*internet: goncharo@eecs.uic.edu*)

ABSTRACT

We present a new method for interpolating between LPC spectra via pole-shifting. This approach solves the problem of real-to-complex and complex-to-real pole transitions by converting to a domain where each pole has a complex conjugate. Desired pole shifts are calculated in the new domain after applying a perceptually-based pole pairing algorithm. Intermediate spectra corresponding to these pole transitions are then optimally approximated using the original number of poles. The resulting interpolated spectral sequence is characterized by approximately linear changes in formant frequencies and bandwidths, and is free of the artifacts that may occur with other LPC spectral parameter interpolation methods.

1. INTRODUCTION

Interpolation of linear predictive coded (LPC) spectra is applied in a variety of speech processing tasks, including speech coding and text-to-speech synthesis. When LPC or cepstral coefficients are interpolated for this purpose there is no guarantee of stability for the filters represented by the intermediate coefficient values. Even when a parameter set used for interpolation guarantees stability there may still be a problem with the naturalness of the spectral sequence (i.e. non-speech-like spectral transitions) as the starting spectrum is incrementally transformed to the desired ending spectrum. Unnatural spectral transitions include the sudden appearance and disappearance of narrowband peaks, peaks fading and appearing at closely spaced frequencies (instead of simply shifting from one frequency to the other), and peak frequency shifts that are highly nonlinear. To avoid these problems during LPC spectral interpolation, we chose to directly manipulate the LPC poles. Shifting poles in the z-plane provides great control over spectral transitions since each pole additively contributes a peak to the log spectrum. In developing a pole shifting method, the following problems were addressed: (1) pairing poles in the starting spectrum with poles in the ending spectrum in a perceptually-meaningful way; (2) defining the desired path for each pole to follow during the interpolation process, and (3) handling the cases of complex-to-real and real-to-complex pole transitions.

2. POLE SHIFTING SCHEME

We have based our algorithm on the assumption that changes in formant frequencies and bandwidths from frame to frame during the interpolation should be linear. Certainly this is true in actual speech over only a short time span, but without additional contextual information this is probably the best assumption that may be made. Given the LPC filter model

$$H(z) = \frac{G}{1 - \sum_{k=1}^{N} \alpha_k z^{-k}} = \frac{G}{\prod_{j=1}^{N} \left(1 - p_j z^{-1}\right)}$$

where G is a gain constant, α_k $\{k = 1, \ldots, N\}$ are the LPC coefficients modeling the effects of vocal tract filtering, and p_j $\{j = 1, \ldots, N\}$ are the corresponding poles of $H(z)$, we may define the frequency and bandwidth of each pole p_i:

$$F_i = \frac{\angle p_i}{2\pi T_s} = \frac{\theta_i}{2\pi T_s}, \qquad BW_i = \frac{-\ln|p_i|}{\pi T_s} = \frac{-\ln r_i}{\pi T_s},$$

with T_S representing the sampling period. These values are those that match the center frequency and bandwidth of the frequency response resulting from a single pole inside the

0-7803-2431-5/95 $4.00

unit circle on the z-plane. To achieve linear changes in F_i and BW_i versus some blending parameter a, define the i-th pole's frequency and bandwidth tracks as follows (a varies from 0 to 1):

$$F_i(a) = (1-a)F_{i0} + aF_{i1}$$
$$BW_i(a) = (1-a)BW_{i0} + aBW_{i1}$$

It is clear that the frequency of p_i linearly changes from F_{i0} to F_{i1}, and that the bandwidth of p_i linearly changes from BW_{i0} to BW_{i1} as a function of a. Henceforth we drop the subscript specifying pole index and refer to an arbitrary pole p_i as p. It may be shown that linear changes in F and BW versus a occur when:

$$p(a) = p_0\left(\frac{p_1}{p_0}\right)^a.$$

As before, the subscript 0 refers to the starting value (a = 0) and 1 to the ending value (a = 1) of each parameter during interpolation. The pole interpolation paths defined by the equation above appear spiral-like in shape.

3. POLE PAIRING PROCEDURE

In order to apply this pole shifting scheme, we found it necessary to establish a pole pairing procedure that relates to both the pole transition paths and some perceptually-meaningful criteria. Based on the distance measure presented in this section, the selected pole pairings are those that result in the minimum cumulative distance between poles. In calculating distances all possible pole pairings must be considered. For N poles this requires that $N!$ pole combinations be evaluated. The problem of a more efficient sub-optimal pole pairing algorithm is left for future work. Based on the derived pole path function $p(a)$, the length of the path taken from p_0 to p_1 is

$$L(p_0, p_1) = \int_0^1 \left|\frac{\partial p(a)}{\partial a}\right| da,$$

which serves as the basis for our distance measure. Using a weighting factor proportional to the perceptual effect of $p(a)$, the distance between the starting and ending pole locations is defined as a weighted calculation of path length:

$$D(p_0, p_1) = \int_0^1 \left| W(p(a))\,\frac{\partial p(a)}{\partial a}\right| da$$

This weighting function was derived based on the power spectrum of a single-pole filter. Because it is possible that a half-power point (-3 dB from the peak) does not exist in this response, we instead define ω_{mid} as the frequency at which the power level is halfway between its minimum and maximum values:

$$\left|H\left(e^{j\omega_{mid}}\right)\right|^2 = \frac{1}{2}\left(\left|H\left(e^{j\omega}\right)\right|^2_{\min} + \left|H\left(e^{j\omega}\right)\right|^2_{\max}\right)$$

For a pole with magnitude r and angle $\theta = 0$ this frequency is found to be:

$$\omega_{mid} = Cos^{-1}\left(\frac{2r}{r^2+1}\right)$$

Because there is a direct relationship between our perceived loudness of a sound and the logarithm of its power level, we are interested in the effect of changes in pole radius and angle on the log spectrum. In this case a convenient way to measure these effects is to calculate the partial derivative of the log spectrum with respect to both ω and r, at $\omega = \omega_{mid}$. The calculation results in measures that are equal:

$$\left.\frac{\partial \ln\left|H\left(e^{j\omega}\right)\right|^2}{\partial \omega}\right|_{\omega=\omega_{mid}} = \left.\frac{\partial \ln\left|H\left(e^{j\omega}\right)\right|^2}{\partial r}\right|_{\omega=\omega_{mid}} = \frac{2r}{1-r^2}$$

Thus small shifts of the pole at $z = r$ in either radial or angular directions have the same perceptual effect at our reference frequency ω_{mid}. We conclude that the weighting function need not discriminate between radial and angular pole movements when applying perceptually-meaningful weighting. Our choice for this weighting function is:

$$W(p(a)) = W(a) = 2r(a)\left[1-r(a)^2\right]^{-1}$$

We take this expression from the sensitivity measures calculated previously; note that the expression assigns infinite weight to poles on the unit circle, and zero weight to poles at the origin. Substituting the weighting function into the distance measure expression yields:

$$D(p_0, p_1) = \int_0^1 \left| \frac{2r(a)}{\left(1 - r(a)^2\right)} \frac{\partial p(a)}{\partial a} \right| da$$

Solving this integral gives us the following:

$$D(p_0, p_1) = \left| \ln\left(\frac{p_1}{p_0}\right) \right| \left\{ \frac{\ln\left(\frac{1-r_0^2}{1-r_1^2}\right)}{\ln\left(\frac{r_1}{r_0}\right)} \right\} \text{ when } r_0 \neq r_1;$$

$$= \left| \ln\left(\frac{p_1}{p_0}\right) \right| \left\{ \frac{2r^2}{1-r^2} \right\} \text{ when } r = r_0 = r_1.$$

4. COMPLEX-CONJUGATE TO REAL POLE MAPPING

A problem inherent to z-plane pole shifting (and previously discussed in [1]) is that of interpolating between real and complex-conjugate pole pairs. Since every LPC coefficient α_k must be real-valued, the roots of the denominator polynomial $A(z)$ must always be real or come in complex-conjugate pairs. Inevitably a set of complex-conjugate poles will be paired with two real poles at different locations in the z-plane, forcing a violation of the complex-conjugate relationship along the interpolation pathway. Our solution to this problem considers the transfer function $H^2(z,a)$, where each pole (whose position is a function of a) retains a complex conjugate partner over its entire interpolation path. To ensure that complex conjugate pole-pairs map to other complex conjugate pole-pairs, the actual pairing is done using "half-plane" transfer functions of the form:

$$\widetilde{H}(z,a) = \frac{G}{\prod_{i=1}^{N}\left(1 - \widetilde{p}_i(a) z^{-1}\right)},$$

where $\widetilde{p}_i(a) = \mathrm{Re}\{p_i(a)\} + j\left|\mathrm{Im}\{p_i(a)\}\right|$.

Once the N poles of $\widetilde{H}(z,0)$ are paired with the N poles of $\widetilde{H}(z,1)$, the desired pole mapping between the $2N$ poles of $H^2(z,0)$ and the $2N$ poles of $H^2(z,1)$ is:

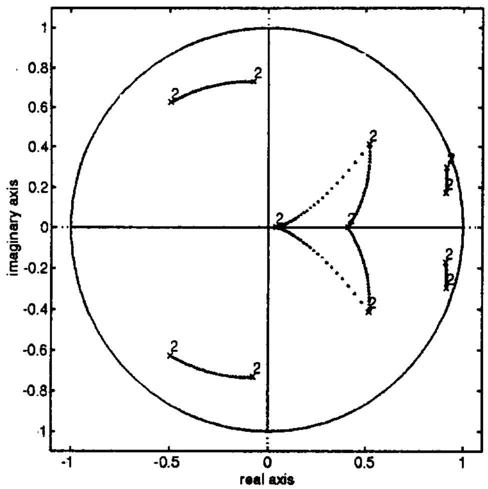
$$\widetilde{p}_i(0) \rightarrow \widetilde{p}_i(1), \quad \widetilde{p}_i^*(0) \rightarrow \widetilde{p}_i^*(1), \qquad i = 1, \ldots, N.$$

If the desired end-result of the interpolation procedure is a sequence of magnitude spectra, they are simply found as $H(z,a) = \sqrt{H^2(z,a)}$. It is more difficult to arrive at the values of N poles $p_i(a)$, $i = 1, \ldots, N$, whose magnitude response best approximates this desired $H(z,a)$. For this we applied the autocorrelation method of calculating LPC coefficients using the Levinson-Durbin recursion, where the autocorrelation function along every step of the interpolation path is found by calculating the inverse FFT of: $H^2(e^{j2\pi k/M})$, $k = 0, \ldots, M-1$.

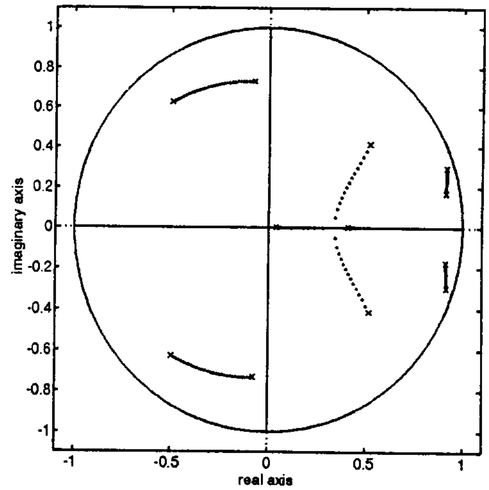
5. RESULTS

Figs. 1 and 2(a) illustrate interpolation of LPC spectra via the methods presented. The pole mapping and shifting procedures for transfer functions $H^2(z,a)$ and the optimal approximation to $\sqrt{H^2(z,a)}$ are plotted in Figs. 1(a) & (b), with the sequence of magnitude spectra resulting from (b) plotted in Fig. 2(a). As a comparison to our methods, the sequence of magnitude spectra resulting from cepstral coefficient interpolation is plotted in Fig. 2(b). While the sequence in Fig 2(b) exhibits unnatural spectral transitions, the sequence in Fig. 2(a) is characterized by approximately linear changes in peak frequencies and bandwidths.

[1] J. M. Turnbull, A. T. Sapeluk and R. I. Damper, "A New Method of Pole-Tracking with Application to Vowel and Semi-vowel Recognition," Proc. IEEE ICASSP 1989, pp. 568-571.

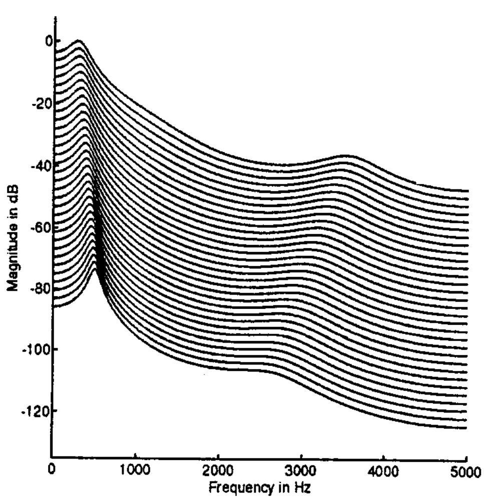


(a) 2N-th order mapping

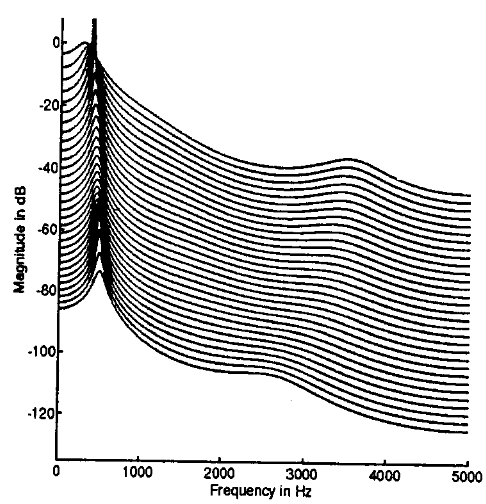


(b) Resulting N-th order mapping

Figure 1: Pole Mapping/Shifting Scheme



(a) Spectral sequence from pole interpolation



(b) Spectral sequence from cepstral coefficient interpolation

Figure 2: LPC Spectra