POLE-ZERO ANALYSIS OF SPEECH BY LINEAR PREDICTIVE CODING

D. Nandagopal*, D.A.H. Johnson* and J.P.T. KoljonenO

INTRODUCTION

The number of applications of the Linear Predictive Coding (LPC) method of processing speech sounds in general, has grown enormously. This is largely due to the time domain nature of the analysis and ease of computation. The majority of its applications have been in the linear all-pole modelling of speech although actual speech production is known to produce zeros. Short-time speech spectral matching by pole-zero modelling has been attempted [Atal and Schroeder, 1978, Makhoul, 1975, Yegnanarayana, 1981].

The cepstral matching criterion, analogous to autocorrelation matching in LPC analysis, has been applied to pole-zero decomposition of short time speech segments [Yegnanarayana, 1981]. In this work, the cepstral coefficients of the model impulse response are equated to the cepstral coefficients of the speech signal up to a specified number determined by the chosen order of the system model. Pole-zero decomposition of speech by LPC has been performed by us as a modification of Yegnanarayana's method [Nandagopal, Johnson and Koljonen, 1986]. We present here a speech processing procedure towards 'WORD' characterization using the pole-zero decomposition of speech based on Linear Predictive Coding.

POLE-ZERO DECOMPOSITION BY LPC

In the modified pole-zero decomposition technique, the cepstral coefficients that are necessary to calculate the group delays are computed from LPC constants by using a large model order and assuming the signal to be minimum phase. This sets all the roots of the model transfer function within the unit circle. The above assumption allows the use of the properties of minimum phase polynomials. If the model transfer function $A(z)$ is an $M$'th order polynomial in $z^{-1}$ then:

$$
\ln[A(z)] = - \sum_{k=1}^{\infty} C_k Z^{-k}
$$

where $C_k$s are the cepstral coefficients.

The above equation defines the relationship between the cepstral coefficients and the roots of the polynomial. If $A(Z)$ is considered to be an inverse filter model of the signal, then the filter coefficients are directly related to the cepstral coefficients. Gray and Markel [1976] have delineated the following equations establishing the relationship between the cepstral coefficients and the LPCs and thus enabling the computation of cepstral coefficients

$$
C_1 = -a_1 : \quad jC_j = -j a_j - \sum_{k}^{j-1} kC_k a^{j-k} \quad \text{for } j = 2, 3, \ldots, M
$$

and

*Dept. of Electrical & Electronic Engineering, University of Melbourne, Australia.
OAcoustics Lab. Helsinki University of Technology, Finland.
\[ jC_j = - \sum_{k=1}^{M} (j-k)C_{j-k}a_k \quad \text{for } j = M + 1, M + 2, \text{ etc.} \quad (3) \]

where

\( C_k \)s are cepstral coefficients \( a_k \)s are LPC coefficients \( M \) is the LPC model order

The minimum phase assumption also allows the use of the relationship between the negative derivative of phase spectrum (group delay) and the cepstral coefficients and hence helps evaluate the group delays from the cepstral coefficients using the following equation

\[ \theta'(w) = \sum_{k=1}^{\infty} kC_k \cos(kw) \quad (4) \]

where \( \theta'(w) \) is the negative derivative of phase spectrum.

[Yegnanarayana, 1981]

The positive and negative peaks in the negative derivative of phase spectrum correspond to complex poles and complex zeros of the signal respectively. Convolution in time domain is equivalent to addition in the cepstral domain. Separation of the positive and negative portions of the group delay enables the computation of cepstral coefficients corresponding to poles and zeros and thus decomposing the signal into its pole part and zero part.

Now, having split the cepstral coefficients of the original signal into pole and zero parts all that remains is to model the pole part of the signal and zero part of the signal individually. Once again, using the relationship between cepstral and LPC coefficients (equations (2), (3) and (4)), the LPCs corresponding to the pole and zero parts of the signal can be computed from the above cepstral coefficients that are already split to represent the pole and zero parts of the signal. The pole and zero spectra are evaluated by considering two all-pole models. Figure (1) describes the complete pole-zero decomposition procedure.

**SPEECH DATA ANALYSIS**

Speech segments used in the present analysis are extracted from the department of electrical engineering's speech data library sampled at 20 KHz. Speech data files corresponding to short words such as 'TWO', 'KID', 'ROW' etc, are resampled digitally at 10 KHz. This enables the LPC programs to be more efficient and use a lower model order. The pole-zero decomposition procedure corresponding to figure 1 is programmed into an LSI 11/23 minicomputer using DAOS software package. The analysis is carried out on data segments of size 1024 samples. Analysis on successive frames of each of the chosen short words are carried out. Averaging in frequency domain is performed to produce a composite spectrum for a complete word.

**POLE-ZERO EVALUATION**

In order to generate a pole-zero table for words, the poles and zeros are calculated from the LPCs evaluated from the cepstral coefficients belonging to pole and Zero parts of the signal. The pole and zero positions are determined by simply factorizing the numerator and denominator LPC polynomials. If \( D(z) \) is the denominator polynomial represented by the LPCs that correspond to the pole-part of the signal, then \( D(z) \) is given by:
\[
D(z) = 1 - a_1 z^{-1} - a_2 z^{-2}, \ldots - a_M z^{-M} \tag{4}
\]

The roots of the polynomial would provide the 'pole' positions. Bairstow's method is used to determine the roots of the above polynomial by iteratively extracting quadratic factors; and solving each factor for its roots. The procedure is repeated for the numerator polynomial to determine the zeros.

RESULTS AND DISCUSSION

In order to justify the use of LPCs from a large order all-pole system, in extracting the cepstral coefficients for a pole-zero system, we have modelled a short speech segment in the frequency domain using cepstral coefficients extracted from the signal through a standard procedure. The same speech segment is then modelled using cepstral coefficients extracted from LPCs (assuming the signal is all-pole initially).

Figure 2 compares the pole-zero spectra obtained using standard cepstral method and LPC method with the data spectrum. Note the negative peak in the figure 2.b matches very well with the data spectrum and demonstrates the efficacy of the LPC method in modelling poles and zeros.

Table 1 compares pole and zero frequencies of our pole and zero models (calculated using the root finding algorithm mentioned above), with corresponding frequencies read from the data spectrum, for the word "TWO". This table confirms the accuracy of our models. Table 2 shows the dominant pole and zero frequencies for the word "TWO" uttered by 2 different speakers and reveals marked similarities. Pole zero tables for each (single syllable) word could be developed after trials with a variety of speakers.

REFERENCES


\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{Pole} & \textbf{Frequency (Hz)} & \textbf{Zero} & \textbf{Frequency (Hz)} & \textbf{Dominant} & \textbf{Dominant} \\
& & & & \textbf{Pole} & \textbf{Position} \\
From Data & From & From & From & Speaker & Speaker \\
Spectrum & LPC & Data & LPC & TM & NR \\
& Model & Spectrum & Model & Speaker & Speaker \\
From Data & From & From & From & TM & NR \\
& Spectrum & LPC & Spectrum & Model & \\
340 & 362 & 1100 & 1119 & .856 & .940 \\
1640 & 1692 & 2500 & 2502 & .821 & .891 \\
2187 & 2148 & & & .909 & .99 \\
\hline
\end{tabular}

Table 1. Pole Zero Frequency Comparison for word "TWO"

\begin{tabular}{|c|c|}
\hline
\textbf{Speaker} & \textbf{TM} \\
\hline
.850 & .902 \\
.990 & .982 \\
\hline
\end{tabular}

Table 2. Pole Zero Positions for word "TWO"
Figure 1. SPEECH SPECTRA BY VARIOUS METHODS

(a) BY STD METHOD

(b) BY LPC METHOD

(c) DATA SPECTRUM