MODIFIED POLE-ZERO DECOMPOSTION OF SPEECH

D. Nandagopal (*), D.A.H. Johnson (*), J.P.T. Koljonen (**)

(*) Department of Electrical and Electronic Engineering, University of Melbourne

(**) Acoustic's Lab
Helsinki University of Tech., Finland.

ABSTRACT - A modified method of pole-zero modelling of speech using cepstral coefficients is described. Cepstral coefficients are extracted from linear prediction coefficients (LPC) by assuming the signal to be minimum phase. Pole-zero decomposition is done by splitting the signal into a pole part , and a zero part, by using the group delay properties of the signal. The pole and Zero parts are then modelled using LPC.

INTRODUCTION

Estimation of the vocal tract characteristics is an important aspect of speech processing. The vocal system, normally, is approximated by a linear system model. Therefore, the characteristics of the vocal tract model depends on the type and order of the linear system model assumed. All-pole modelling based on linear predictive coding (LPC) has been by far the most popular technique used for speech analysis. The major short- coming of this technique is its inability to model the zeros, which are known to be produced by the vocal tract.

Short-time speech spectral matching by pole-zero modelling has been attempted [Atal and Schroeder 1978; Makhoul, 1975, Yegnanarayana, 1981]. In these, the speech models consisted of both poles and zeros in their transfer function. Yegnanarayana's pole-zero decomposition technique adopted cepstral matching criterion analogous to autocorrelation matching in LPC analysis [Yegnanarayana, 1981]. The ceptstral coefficients of the model impulse response are equated to the cepstral coefficients of the signal up to a specified number determined by the chosen order of the system model. We present in this paper, an extension of Yegnanarayana's pole-zero decomposition method. Basically, the pole-zero decomposition is done by splitting the signal into a pole part and a zero part (by using the group delay properties of the signal) and both are individually modelled using LPC. The cepstral coefficients that are necessary to calculate the group delays are extracted from LPC coefficients by assuming the signal to be minimum phase. This implies that the poles and zeros of the model must lie within the unit circle in the Z-plane. This assumption would then allow the use of the properties of cepstra from minimum phase polynomials.

MINIMUM PHASE CORRESPONDENT.

Properties of cepstra from minimum phase signals have been documented [Oppenheim and Schafer, 1975; Gray and Markel, 1976]. Group delay properties of stable all-pole systems have been reported [Yegnanarayana, 1978]. The spectra of the minimum phase correspondent and the actual signal are identical by definition. Therefore we can (i) relate the LPCs with cepstral coefficients (ii) use the relationship between the group delay (negative derivative of phase spectrum) and cepstral coefficients as given by the following equations:

$$C_1 = a_1 \qquad (1)$$

$$j\, a_j = -j\, a_j - \sum_{k=1}^{j-1} K\, C_k\, a_{j-k} \qquad (2)$$

for $j = 2, 3, \ldots M$

where C_k's are cepstral coefficients
a_k's are LPC coefficients
M is the LPC model order.
[Gray & Markel, 1976]

$$\theta'(\omega) = \sum k\, c(k) \cos k\omega \qquad (3)$$

where $\theta'(\omega)$ is the group delay or negative derivative of phase spectrum [Yegnanarayana, 1981)

POLE-ZERO DECOMPOSITION

Our pole-zero decompostion procedure involves the extensive use of the equations (1), (2) and (3) as described above. Figure 1 delineates the pole-zero decomposition processing steps. This procedure is similar to Yegnanarayana's, except that the cepstral coefficients are extracted from equations (1) and (2). The group delays are then calculated directly from the ceptstral coefficients using equation (3). As the signal is assumed to be minimum phase, the negative derivative of phase spectrum would have positive peaks corresponding to the complex poles, and negative peaks corresponding to complex zeros of the given signal. Therefore, the next logical step is to separate the positive and negative portions of the group delay. The positive part is the group delay of the pole spectrum and negative part is the group delay of the zero spectrum. Convolution in time domain is equivalent to addition in the cepstral domain. Since the cepstral coefficients correspond to the log spectrum of a pole-zero system, pole-zero decomposition is achieved once the cepstral coefficients are split into a pole part and a zero part. Therefore the next step towards pole-zero decomposition is to retrieve the cepstral coefficients corresponding to the pole-part and the zero part of the data spectrum (according to their respective group delays). This is achieved by the use of equation (3) again, i.e. by performing a reverse Fourier transform and weighting the coefficients appropriately. Now, the pole spectrum can be computed by using the equations (1) and (2) to translate the pole and zero part of cepstral coefficients into LPC coefficients. This procedure is identical for pole and zero parts as indicated in figure 1. Finally the pole-zero spectrum is obtained by the addition of pole and zero spectra.

RESULTS

Speech data sampled at 20 KHZ are used in the present analysis. A short speech segment of 256 samples was used. The pole-zero decomposition algorithms corresponding to figure 1 were programmed into a PDP11/23 mini computer using DAOS software package. Cepstral coefficients are extracted from a speech data segment using LPC coefficients. Cepstral coefficients are the Fourier Coefficients of the log spectrum of speech data. Figure 2 indicates the group delays calculated from the Cepstral coefficients using equation (3). Note the presence of positive and negative peaks and the symmetry. The positive and negative components of the group delay are simply split by rectifying the group delay data. This is indicated in figure 3. The pole and zero spectra are evaluated as described in the previous section and shown in figure 4. Note the dips in the zero spectrum indicating the presence of zeros. Figure 5 compares the data spectrum with the pole zero spectrum for the same data segment. It is evident that the pole-zero spectrum of the model approximates the data spectrum more closely.

DISCUSSION AND CONCLUSION

The main problems in pole-zero modelling of speech are:

1. The harmonic structure of speech spectra, which can cause the pole zero algorithm to erroneously detect zeros between harmonic peaks.
2. Noise between the harmonic peaks, which affects cepstral smoothing algorithms.
3. Accuracy loss due to the dynamic range of the magnitude spectrum.

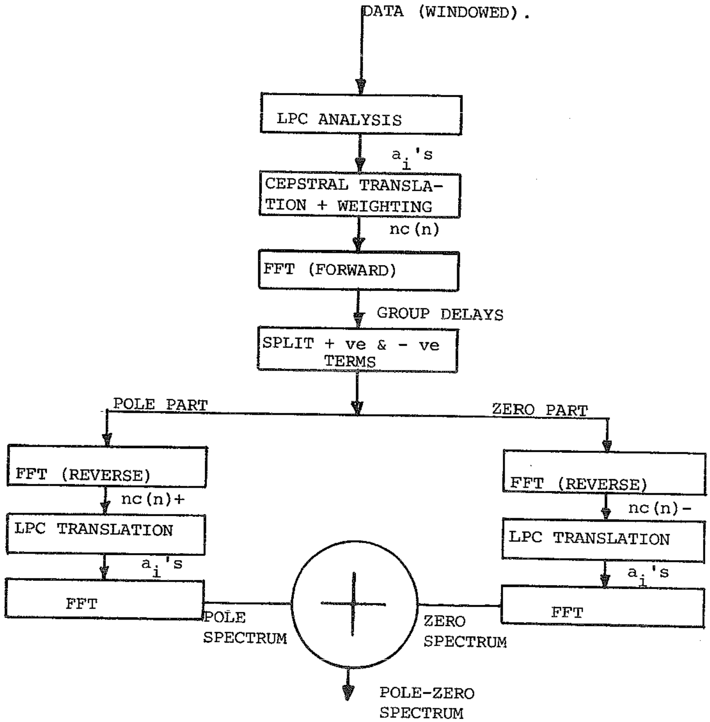

Figure 1 POLE-ZERO DECOMPOSITION PROCEDURE

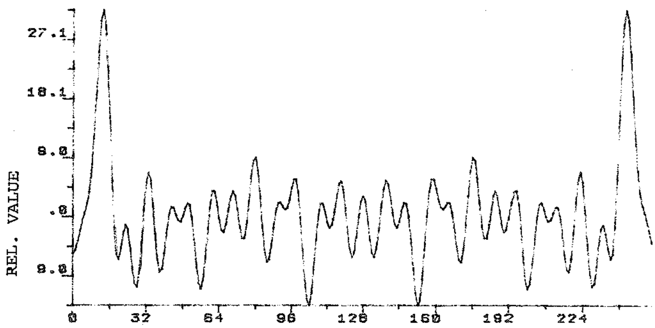

Figure 2 GROUP DELAY

Problems 1 and 3 can be overcome using cepstrum analysis techniques. However, it doesn't handle effectively the noise problem due to the uniform weighting property of the cepstral method (noise corrupted spectral valleys are weighted as much as the "clean" peaks). In our method, this particular problem is avoided by the use of LPC analysis as it lays emphasis on the spectral peaks. Also estimation of cepstral coefficients is done by the use of simple relationships with the LPCs. The LPCs are obtained using simple and rapid time domain operations. We have presented here our preliminary work leading into the development of a new technique for pole-zero modelling of speech. Although the present procedure is computation intensive (typically 3 minutes for 256 samples, when run under DAOS), the results are encouraging and there is scope for further improvement.

REFERENCES

Atal, B.S. and Schroeder, M.R. (1978), "Linear Prediction Analysis of Speech based on a pole-zero representation", J. Acoust. Soc. Am. Vol. 64, No. 5 , 1310-1318.

Gray, A.H. and Markel, J.D. (1976), "Distance Measures for Speech Processing" IEEE Trans. ASSP, vol. 24, No. 5, 380-391.

Oppenheim, A.V. and Schafer, R.W. (1975), "Digital Signal Processing," Prentice-Hall Englewood Cliffs, New Jersey.

Makhoul, J.I. (1975), "Linear Prediction : A tutorial Review", Proc. IEEE, Vol 63, No. 4, 561-580.

Yegnanarayana, B. (1981), "Speech Analysis by Pole-zero decomposition of Short-time Spectra," Signal Processing, Vol 3, No 1, 5-17.

Yegnanarayana, B. (1978), "Formant extraction from linear-prediction phase spectra", J. Acoust. Soc. Am. Vol 63, No. 5, 1638-1640.

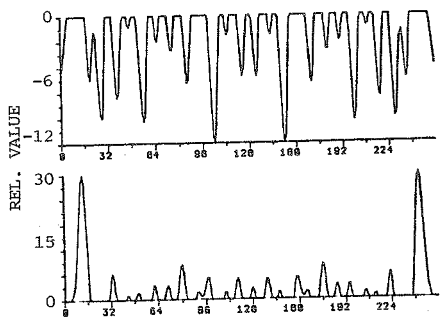

Fig. 3. POSITIVE AND NEGATIVE GROUP DELAYS.

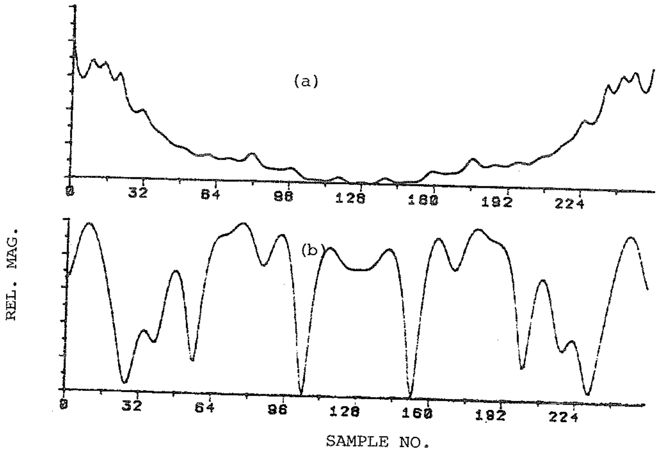

Fig. 5 (a) POLE SPECTRUM, (b)ZERO SPECTRUM

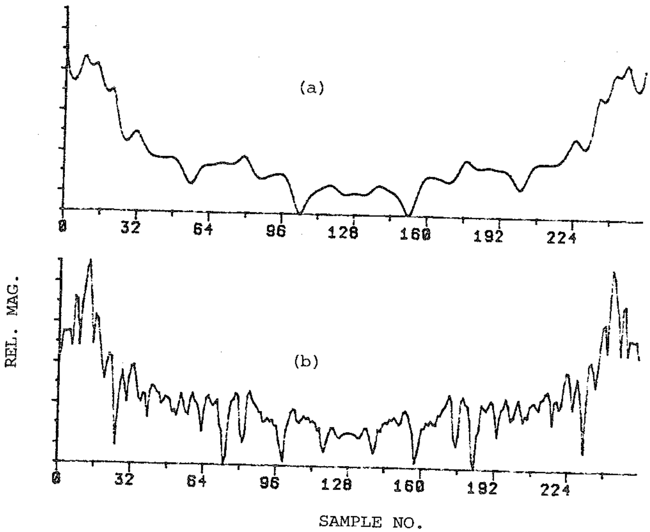

Fig. 5 (a) P-Z MODEL SPECTRUM, (b) DATA SPECTRUM