A POLE-ZERO MODEL FOR CEPSTRALLY SMOOTHED SPEECH SPECTRA

B. Yegnanarayana†

ISRO SATELLITE CENTRE
Peenya, BANGALORE-560 058, INDIA

ABSTRACT

In this paper a new method for representation of cepstrally smoothed speech spectra by a pole-zero model is presented. In this method the cepstrally smoothed log spectrum is split into two parts, one corresponding to the response of the numerator polynomial of the model transfer function and the other part to the response of the denominator polynomial of the model transfer function. The decomposition is achieved by using the properties of the derivative of phase spectra of minimum phase signals. The inverse of each of these responses is approximated by a small number of autoregressive coefficients. The method is illustrated with several examples of speech spectra. The residual from the inverse pole-zero model system can be used to obtain information about the excitation signal. The technique proposed in this paper can be used to represent any arbitrary smoothed log spectrum by a pole-zero model of appropriate order.

I. INTRODUCTION

An important problem in speech analysis is the estimation of the characteristics of the vocal tract system and the excitation source from speech signal. Due to nonstationary nature of speech the analysis is performed by assuming stationarity over short durations (20-40 m sec.) of the signal. The analysis is performed by approximating the vocal tract system by a linear system model and estimating the parameters of the model by adopting an error criterion. The excitation information is derived by passing the signal through the inverse of the model system. The accuracy of analysis depends on the accuracy of representation of the signal characteristics by the model system. In general the model system is derived so as to represent the smoothed short-time spectrum of speech. The fine structure of the spectrum is used to derive the excitation information.

A linear system model consisting of both poles and zeros in its transfer function is required to represent the characteristics of peaks and valleys in the smoothed short-time spectrum of speech. Approximating speech spectra by pole-zero models and estimating the parameters of such models has recently been the subject of active research [1], [2]. We present in this paper a new method for determining the pole and zero parameters of the model. The basic idea is to split the smoothed short-time spectrum into two component responses, each of which can be approximated by a small number of parameters. The smoothed log spectrum is obtained using the first few (20-40 at a sampling rate of 10 kHz) cepstral coefficients, which are the Fourier coefficients of the log spectrum of speech data. Convolution in time domain is equivalent to addition in the cepstral domain. If the cepstral coefficients correspond to the log spectrum of a pole-zero system, then a pole-zero deconvolution can be achieved if the coefficients are split into pole part and zero part. We show that such a splitting can be accomplished by using the properties of the derivative of phase function of a minimum phase signal [3]. The resulting component responses can be approximated by a small number of parameters through autoregressive modelling [2].

In Sec.II & III the problem of pole-zero estimation and the underlying principle of the proposed technique for pole-zero decomposition is presented. An algorithm for pole-zero decomposition of speech spectra is presented in Sec. IV. Some examples of pole-zero decomposition of speech spectra are discussed in Sec.V. Effects of various analysis parameters on the accuracy of the resulting pole-zero model are also discussed.

II. PROPERTIES OF THE DERIVATIVE OF PHASE SPECTRUM

In this section the problem and the underlying principle of the proposed method for solving the problem are discussed.

<u>The Problem:</u>

Let us represent a pole-zero model by

$$H(z) = G\ N(z)/D(z), \qquad (1)$$

where G is a gain term,

$$N(z) = 1 + \sum_{k=1}^{M_2} a^-(k)\ z^{-k} \qquad (2)$$

and

$$D(z) = 1 + \sum_{k=1}^{M_1} a^+(k)\ z^{-k} \qquad (3)$$

CH1610-5/81/0000-0347 $00.75

The problem is to determine the parameters of $H(z)$ such that the frequency response of the model matches the smoothed spectrum of a segment of speech data $x(n)$.

Basis for Pole-zero Decomposition

If all the poles and zeros of the filter $H(z)$ lie within the unit circle in the z-plane, then the filter is called a minimum phase filter. Properties of NDPS* of minimum phase filters are described in [4]. In particular, it can be shown that significant values of NDPS for a minimum phase all-pole filter are positive and for a minimumphase all-zero filter are negative. The significant values of NDPS are confined to regions near frequencies of resonance for an all-pole filter and to regions near frequencies of antiresonance for an all-zero filter. The values of NDPS near zero or folding frequency are contributed by real poles and zeros. The height of the peak at a resonance is inversely proportional to its bandwidth. The contributions of several filters in cascade are additive in the NDPS domain. Using these properties of the negative derivative of minimum phase spectrum it is possible to separate the significant contributions of poles and zeros in the combined NDPS response of a pole-zero filter by considering the positive and negative portions respectively.

III. POLE-ZERO ANALYSIS

Relation Between Derivative of Phase Spectrum and Cepstral Coefficients:

Let $V(\omega)$ be the Fourier transform of the minimum phase correspondent of a given signal. For uniformly sampled discrete signals the Fourier transform is periodic in with period 2π. Since all the poles and zeros of $V(\omega)$ lie within the unit circle in the z-plane, $\ln V(\omega)$ can be expressed in Fourier series expansion as follows [4]:

$$\ln V(\omega) = c(0)/2 + \sum_{k=1}^{\infty} c(k)\, e^{-jk\omega} \qquad (4)$$

where $\{c(k)\}$ are called cepstral coefficients. Writing

$$V(\omega) = |V(\omega)|\, e^{-j\theta_V(\omega)}, \qquad (5)$$

we get the NDPS of $V(\omega)$ as

$$\theta_V'(\omega) = \sum_{k=1}^{\infty} k\, c(k) \cos k\omega \qquad (6)$$

Pole-zero Decomposition

Separating the positive and negative parts of $\theta_V'(\omega)$, we get the approximate NDPS of the pole and zero components of the filter respectively.

$$\theta_V'(\omega) = [\theta_V'(\omega)]^+ + [\theta_V'(\omega)]^- \qquad (7)$$

where

$$[\theta_V'(\omega)]^+ = \theta_V'(\omega), \text{ for } \theta_V'(\omega) \geq 0 \text{ (NDPS of pole part)}$$
$$= 0, \text{ for } \theta_V'(\omega) < 0 \qquad (8)$$

and

$$[\theta_V'(\omega)]^- = \theta_V'(\omega), \text{ for } \theta_V'(\omega) < 0 \text{ (NDPS of zero part)}$$
$$= 0, \text{ for } \theta_V'(\omega) \geq 0 . \qquad (9)$$

Expressing each of the NDPS responses separately in Fourier series yields the cepstral coefficients $\{c^+(k)\}$ and $\{c^-(k)\}$ corresponding to pole and zero spectrum respectively. That is

$$[\theta_V'(\omega)]^+ = C + \sum_{k=1}^{\infty} k\, c^+(k) \cos k\omega \qquad (10)$$

and

$$[\theta_V'(\omega)]^- = -C + \sum_{k=1}^{\infty} k\, c^-(k) \cos k\omega, \qquad (11)$$

where C is the average value, and $c^+(k) + c^-(k) = c(k)$. From $\{c^+(k)\}$ and $\{c^-(k)\}$ the pole spectrum and zero spectrum can be computed through Fourier cosine transform and exponentiation.

Derivation of Model Parameters

We now describe a method of obtaining the parameters of a pole-zero model that represents the cepstrally smoothed spectrum of a signal. For cepstrally smoothed spectrum $\theta_V'(\omega)$ in equation (6) is given by

$$\theta_V'(\omega) = \sum_{k=1}^{M} k\, c(k) \cos k\omega \qquad (12)$$

where M is the length of the cepstral window used for smoothing. Let the linear system given in (1) represent the pole-zero model we are trying to determine. Since the poles and zeros of $H(z)$ lie within the unit circle in z-plane, the numerator and the denominator polynomials can be considered as two inverse filters of linear prediction analysis. The $\{c^+(k)\}$ and $\{c^-(k)\}$ represent the cepstral coefficients corresponding to the two component spectra i.e. the pole part and the zero part respectively. From $\{c^+(k)\}$ and $\{c^-(k)\}$ the pole spectrum $P(\omega)$ and the zero spectrum $Z(\omega)$ can be derived through Fourier transform and exponentation.

The spectra $P(\omega)$ and $1/Z(\omega)$ are approximated using autoregressive modelling. The autocorrelation coefficients $\{R^+(k)\}$ and $\{R^-(k)\}$ corresponding to $P(\omega)$ and $1/Z(\omega)$ are given by the relations:

$$P(\omega) = R^+(0) + 2\sum_{k=1}^{\infty} R^+(k) \cos k\omega \qquad (13)$$

and

$$1/Z(\omega) = R^-(0) + 2\sum_{k=1}^{\infty} R^-(k) \cos k\omega\ . \qquad (14)$$

$\{R^+(k)\}$ and $\{R^-(k)\}$ can be obtained from $P(\omega)$ and $1/Z(\omega)$ respectively using inverse Fourier transform. The autoregressive coefficients $\{a^+(k)\}$ and $\{a^-(k)\}$ are derived from $\{R^+(k)\}$ and $\{R^-(k)\}$ respectively using Levinson's algorithm for solving autocorrelation normal equations [2]. The gain term G in (1) is given by $G = \exp[c(0)/2]$.

IV. POLE-ZERO DECOMPOSITION OF SPEECH SPECTRA

So far the general theoritical basis for pole-zero decomposition has been discussed. In this section we present an algorithm for computing the parameters of the model with specific reference to speech signals.

In this paper we consider speech signals sampled at 10 kHz. A segment of 512 samples is used for analysis. The data is multiplied with a Hamming window before computing the spectrum. The derivative of the phase spectrum is computed from the first M cepstral coefficients. The choice of M depends on the accuracy of representation. The effect of the parameters M, M_1 and M_2 on the resulting model are discussed in Sec. V. All the discrete Fourier transformers (DFT) reported in this paper are computed using a 512-point FFT.

The algorithm for pole-zero decomposition and the determination of model parameters are summarized below:

1. Select a segment of speech data of width 512 samples.
2. Multiply the data with Hamming window.
3. Compute 512-point DFT.
4. Compute log spectrum.
5. Compute cepstral coefficients $\{c(k)\}$ using inverse DFT.
6. Multiply the cepstral coefficients with a rectangular window of width M + 1 samples.
7. Compute NDPS from the windowed cepstral coefficients using DFT.
8. Split the NDPS into positive and negative portions.
9. Find the cepstral coefficients $\{c^+(k)\}$ and $\{c^-(k)\}$ using IDFT.
10. Compute the component log spectra using DFT.
11. Compute the pole spectrum and zero spectrum using exponentation.
12. Compute the autocorrelation coefficients $\{R^+(k)\}$ and $\{R^-(k)\}$ using IDFT.
13. Solve for $\{a^+(k)\}$ and $\{a^-(k)\}$ from the autocorrelation normal equations.
14. $\{a^+(k)\}$ and $\{a^-(k)\}$ determine the response of the model system.

V. RESULTS AND DISCUSSION

In this section we consider some examples of speech spectra to illustrate the application of the proposed method.

The value of M determines the width of the window in the cepstral domain used for computation of the derivative of phase spectrum. It is clear that a larger value of M produces a derivative of phase spectrum with increased resolution for peaks and valleys in the smoothed spectrum.
For a 51.2 msec segment of voiced speech the pole spectrum $P(\omega)$, the zero spectrum $Z(\omega)$ and the cepstrally smoothed spectrum $P(\omega)\ Z(\omega)$ for M = 20 are shown in Fig.1. The figure shows the complementary nature of pole and zero spectra. The pole spectrum has narrow peaks and broad valleys, whereas the zero spectrum has broad peaks and narrow valleys.

The pole spectrum and the inverse of zero spectrum are approximated by autoregressive models of order M_1 and M_2 respectively. The pole-zero model spectra for different values of M_1 and M_2 show the nature of approximation to the cepstrally smoothed spectrum. This is illustrated in Fig.2. Depending on the spectral fit required the values of M_1 and M_2 may be appropriately chosen.

VI. CONCLUSIONS

A new technique for representing the smoothed short-time spectrum of speech by a pole-zero model has been presented. The order of the model can be chosen depending on the accuracy of representation of peaks and valleys in the smoothed spectrum. It appears that any smooth magnitude frequency response can be realised by a pole-zero system of a low order using the technique presented in this paper.

REFERENCES

1. B.S.Atal and M.R.Schroeder, "Linear prediction analysis of speech based on a pole-zero representation", J.Acoust Soc.Amer., vol.64, No.5, pp.1310-1318, 1978.
2. J.I.Makhoul, "Linear prediction: A tutorial review", Proc. IEEE Vol.63 pp.561-580, April 1975.
3. B.Yegnanarayana, "Formant extraction from linear prediction phase spectra", J.Acoust. Soc. Amer., vol.63, pp.1638-1640, 1978.
4. A.V.Oppenheim and R.W.Schafer, Digital Signal Processing, Prentice - Hall, 1975, Ch.10.

† Present Address:
Professor, Computer Centre, Indian Institute of Technology, Madras-600 036, INDIA.

* NDPS: Negative derivation of phase spectrum.

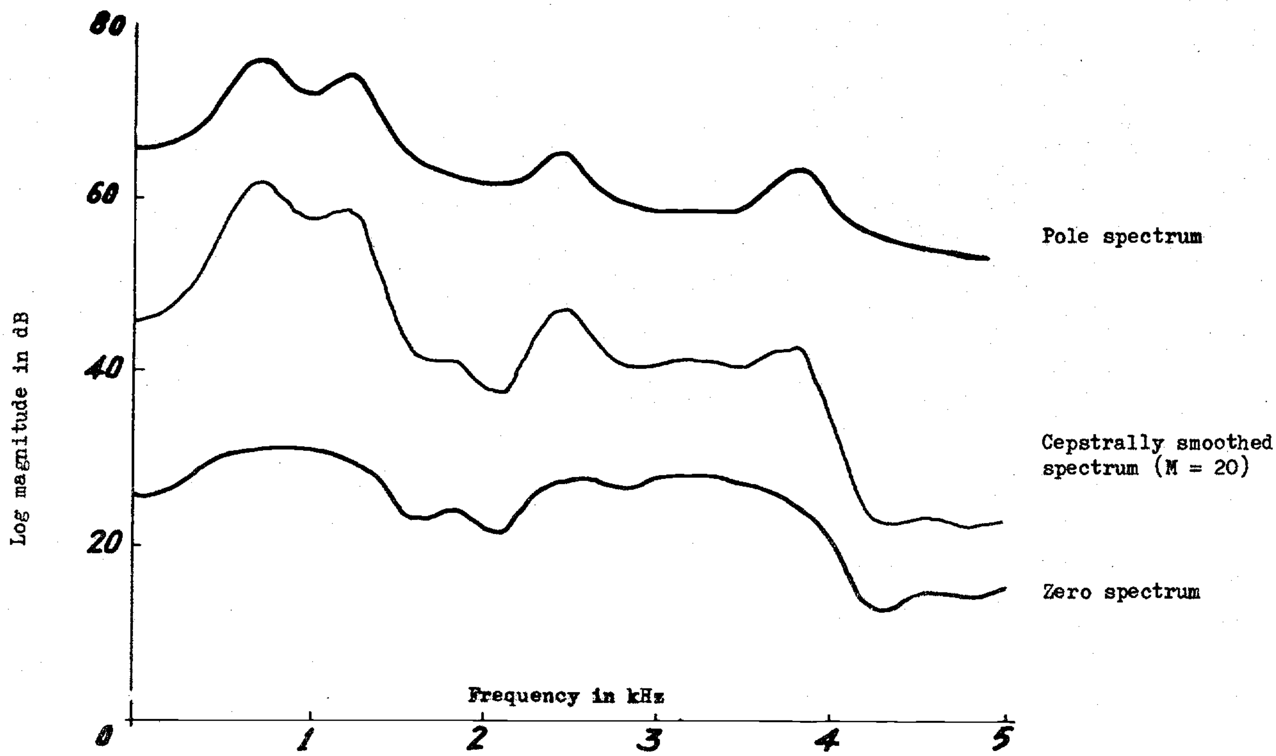

Fig.1. Component Spectra and cepstrally smoothed spectrum for a segment of voiced speech.

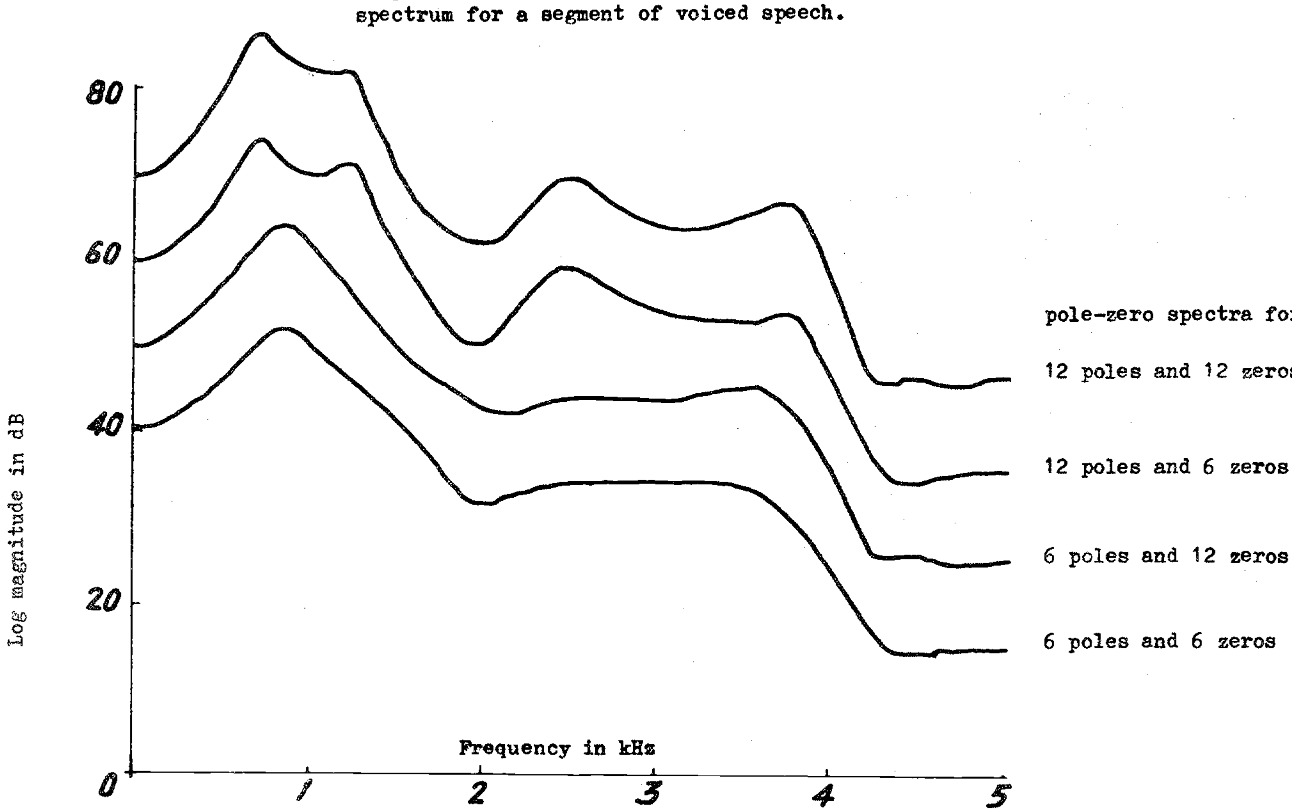

Fig.2. Pole-zero spectra for different values of M_1 and M_2