Signal Processing 3 (1981) 5–17

SPEECH ANALYSIS BY POLE-ZERO DECOMPOSITION OF SHORT-TIME SPECTRA

B. YEGNANARAYANA
Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA 15213, USA
(Present address: *Computer Center, Indian Institute of Technology, Madras-600036, India*)

Received 28 January 1980
Revised 2 June 1980

Abstract. A new method for representation of speech spectra based on a pole-zero decomposition technique is proposed in this paper. In this method the parameters of a pole-zero model for the smoothed short-time spectrum of speech are determined by adopting a cepstral matching criterion. The cepstral coefficients of the impulse response of the model are equal to the cepstral coefficients of the signal up to a specified number which determine the order of the model system. This is analogous to autocorrelation matching in linear prediction analysis. It is shown that the model spectrum represents both peaks and valleys of the smoothed spectrum equally well, unlike the all pole model of linear prediction analysis where only the peaks are well represented. The pole and zero parameters are derived in an identical manner by approximately deconvolving the pole and zero contributions in the cepstral domain. The residual from the inverse pole-zero system can be used to obtain information about the excitation signal.

Zusammenfassung. Es wird eine neue Methode zur spektralen Darstellung von Sprachsignalen vorgeschlagen, die auf der Zerlegung der Spektralfunktion in Pole und Nullstellen basiert. Hierbei werden die Parameter eines Pol-Nullstellenmodells für das geglättete Kurzzeitspektrum von Sprache bestimmt, indem ein Ähnlichkeitskriterium im Cepstrum-Bereich aufgestellt wird. Bis zu einer bestimmten oberen Grenze, die den Grad des Systems bestimmt, sind die Cepstrum-Koeffizienten der Impulsantwort des Modells gleich den Cepstrum-Koeffizienten des Signals. Dies entspricht dem Verhalten der Autokorrelationsmethode der linearen Prädiktion. Mit der beschriebenen Methode werden die Maxima wie auch die Minima des geglätteten Spektrums gleichermassen gut dargestellt; dies im Gegensatz zur linearen Prädiktion, die lediglich die Maxima gut beschreibt. Pole und Nullstellen werden auf gleiche Weise ermittelt, indem ihr Beitrag zur Spektralfunktion im Cepstrum-Bereich mit Hilfe der inversen Faltung näherungsweise getrennt wird. Das Ausgangssignal des aus den Parametern gebildeten inversen Filters vermittelt eine genaue Rekonstruktion des Anregungssignals; insbesondere ist es möglich, daraus den Zeitpunkt des Öffnens und Schliessens der Glottis zu bestimmen.

Résumé. Une nouvelle méthode de représentation des spectres de parole, fondée sur une technique de décomposition en pôles et zéros, est proposée. Dans cette méthode, les paramètres d'un modèle comprenant des pôles et des zéros du spectre à court terme lissé de la parole sont déterminés en utilisant un critère d'ajustement "cepstral". Les coefficients "cepstraux" de la résponse impulsionnelle du modèle sont jusqu'à un ordre donné égaux aux coefficients "cepstraux" du signal. Cet ordre détermine celui du modèle. Ceci est analogue à l'ajustement des valeurs de l'autocorrélation en prédiction linéaire. On montre que le spectre du modèle représente les pics et les vallées du spectre lissé avec la même précision alors que le modèle autorégressif de la prédiction linéaire ne représente bien que les pics. Les paramètres du numérateur et du dénominateur du modèle sont calculés de manière identique en effectuant une déconvolution approximative de la contribution des pôles et des zéros dans le domaine "cepstral". Le résidu du filtrage inverse par pôles et zéros donne une information précise sur le signal d'excitation. En particulier, on peut en déduire la forme ds ondes glottiques avec les périodes d'ouverture et de fermeture.

Keywords. Pole-zero decomposition, minimum phase signal, cepstral coefficients, negative derivative of phase spectrum, linear prediction, pole-zero model, ceptral matching, autocorrelation matching, pole-zero spectrum, quefrency.

1. 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 msec) of the signal. The analysis is performed by approximating the vocal 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 method for determining simultaneously the pole and zero parameters of a pole-zero model. The model parameters are determined by adopting the criterion of cepstral matching. Cepstral coefficients are the Fourier coefficients of the log spectrum of speech data. The first few (20–40 at a sampling rate of 10 kHz) coefficients are normally used to represent the smoothed spectral characteristics and this is the basis for homomorphic deconvolution [3]. 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 made approximately by using the properties of the derivative of phase function of a minimum phase signal [4]. It is possible to derive a pole-zero model from the deconvolved coefficients such that the cepstral coefficients of the model match those of the given data up to a specified number. The cepstral coefficients beyond the specified number are uniquely determined by the model system. In this respect the model spectrum is different from a cepstrally smoothed spectrum where the cepstral coefficients beyond the specified number are set to zero. We show that the criterion of matching cepstral coefficients for determining a pole-zero model is analogous to the criterion of matching autocorrelation coefficients for determining an all pole model. However, since log magnitude spectrum provides a better representation of spectral information than magnitude spectrum itself, due to large dynamic range of the spectrum, matching cepstral coefficients should give a better spectral modelling than matching autocorrelation coefficients.

In Section 2 the problem of pole-zero estimation and the underlying principle of the proposed technique are discussed. In Section 3 the technique for pole-zero decomposition is presented. An algorithm for pole-zero decomposition of speech spectra is presented in Section 4. Several examples of pole-zero decomposition of speech spectra are discussed in Section 5. Effects of various analysis parameters on the accuracy of the resulting pole-zero model are also discussed. Some issues presently under investigation are cited in Section 6.

2. 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.

2.1. The problem

Let us represent a pole-zero model by

$$H(z) = GN(z)/D(z) \qquad (1)$$

where G is a gain term,

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

and

$$D(z) = 1 + \sum_{k=1}^{M} a^{+}(k) z^{-k}. \tag{3}$$

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 $x(n)$. For the system $H(z)$ to be stable all the roots of the denominator polynomial $D(z)$, called poles, must lie within the unit circle in the z-plane. If the roots of the numerator polynomial, called zeros, also lie within the unit circle in the z-plane, then the impulse response of $H(z)$ is called a minimum phase signal. An important property of minimum phase signals is that the magnitude and phase responses are related through Hilbert transformation [3].

2.2. *Basis for pole-zero decomposition*

Since the objective in the present problem is to determine a pole-zero model for a signal spectrum, it is sufficient to consider the minimum phase correspondent of the given signal. The spectra of the minimum phase correspondent and the original signal are identical by definition. Properties of minimum phase signals have been extensively studied [3, 5]. In particular, all the poles and the zeros of a minimum phase signal lie within the unit circle in the z-plane.

Properties of the *derivative* of phase spectrum (DPS) of a stable all-pole system have recently been reported by the author [4] in the context of formant extraction using linear prediction coefficients (LPCs). A stable all-pole system can be represented as a cascade of first order sections with real poles and second order sections with complex conjugate poles. In this paper we consider the negative derivative of phase spectra (NDPS) throughout, for convenience. The NDPS of a typical first order filter (real pole) is given by

$$\Theta_1'(\omega) = \gamma/(\omega^2 + \gamma^2) \tag{4}$$

where γ is the corner frequency. The NDPS of a typical second order filter (resonator) is given by

$$\Theta_2'(\omega) = \frac{2\alpha(\alpha^2 + \beta^2 + \omega^2)}{(\alpha^2 + \beta^2 - \omega^2)^2 + 4\omega^2\alpha^2} \tag{5}$$

where α and β are the half power bandwidth and resonance frequency of the filter, resp. These equations are derived in [4]. In general $\beta^2 \gg \alpha^2$. The NDPS of the overall filter, denoted by $\Theta'(\omega)$, is a summation of the terms of the type given in (4) and (5). Some important properties of $\Theta'(\omega)$ are:

(1) $\Theta_1'(\omega)$ is a monotonically decreasing function of ω.

(2) At low frequencies $\Theta_1'(\omega) \cong 1/\gamma$.

(3) At high frequencies $\Theta_1'(\omega) \cong \gamma/\omega^2$.

(4) $\Theta_2'(\omega)$ is approximately proportional to the squared magnitude response of the filter around the resonance frequency.

(5) At low frequencies $\Theta_2'(\omega) \cong 2\alpha/\beta^2$, which is a small constant quantity.

(6) At high frequencies $\Theta_2'(\omega) \cong 2\alpha/\omega^2$.

It is interesting to note that if the corner frequency γ is large, then $\Theta_1'(\omega)$ will be small for all ω. On the other hand if γ is small, then the large values of $\Theta_1'(\omega)$ are confined to frequencies close to the origin. As a result of the properties (1), (2) and (3) real poles will have negligible effect on the peak structure of $\Theta'(\omega)$ caused by resonances. The properties (4), (5) and (6) show that in $\Theta'(\omega)$ there is negligible effect of one resonance peak on the other.

It is easy to visualize a similar behaviour for real and complex conjugate zeros in their NDPS plots. The only difference is that the NDPS for zeros will have a sign opposite to that for poles. Specifically $\Theta'(\omega)$ will have a positive peak due to a complex conjugate pole pair and a negative peak due to a complex conjugate zero pair. These simple but powerful properties of the derivative of phase spectrum are shown to accomplish the pole-zero decomposition discussed in the next section.

In Fig. 1 the NDPS plots for a first order filter and a second order pole filter are shown. It is clear from the figure that significant values of $\Theta'(\omega)$ are confined to frequencies near the origin for real

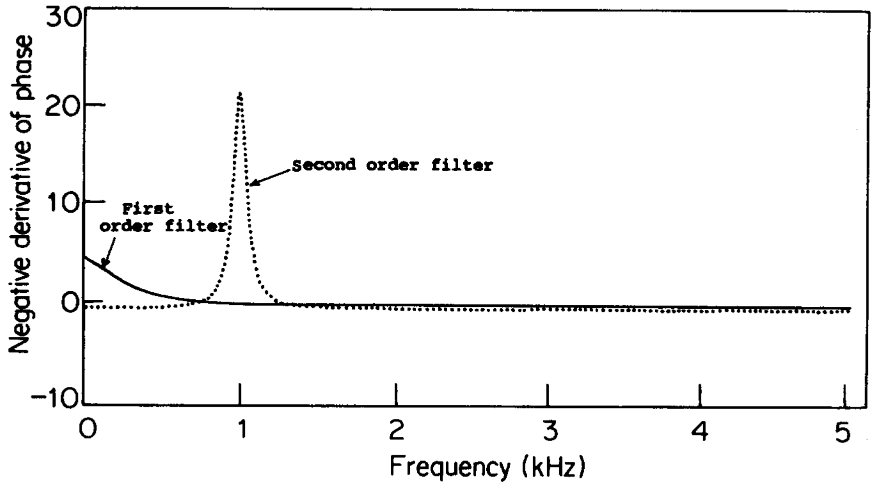

Fig. 1. Negative derivative of phase spectra for typical all pole filters. (a) Solid curve: First order filter $[H(z)=1/(1-0.85z^{-1})]$. (b) Dotted curve: Second order filter $[H(z)=1/(1-1.57z^{-1}+0.94z^{-2})]$.

poles and to frequencies near the resonance frequency for complex conjugate poles.

3. Pole-zero analysis

3.1. 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 ω with period 2π. Since all the poles and zeros of $V(\omega)$ lie within the unit circle in the z-plane [3], $\ln V(\omega)$ can be expressed in Fourier series expansion as follows:

$$\ln V(\omega)=\tfrac{1}{2}c(0)+\sum_{k=1}^{\infty} c(k)\,\mathrm{e}^{-jk\omega} \tag{6}$$

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

$$V(\omega)=|V(\omega)|\mathrm{e}^{-j\Theta_V(\omega)} \tag{7}$$

we get the real and imaginary parts of $\ln V(\omega)$ as

$$\ln|V(\omega)|=\tfrac{1}{2}c(0)+\sum_{k=1}^{\infty} c(k)\cos k\omega \quad \text{(real part)} \tag{8}$$

and

$$\Theta_V(\omega)+2\lambda\pi=\sum_{k=1}^{\infty} c(k)\sin k\omega \quad \text{(imaginary part)} \tag{9}$$

where λ is an integer. Notice that $\Theta_V(\omega)$ represents the negative phase spectrum of a minimum phase signal. Taking the derivative of $\Theta_V(\omega)$, we get

$$\Theta'_V(\omega)=\sum_{k=1}^{\infty} kc(k)\cos k\omega. \tag{10}$$

3.2. Pole-zero decomposition

$\Theta'_V(\omega)$ is the negative derivative of phase spectrum of a minimum phase signal whose properties were discussed in Section 2. In particular, the complex poles of $V(\omega)$ produce positive peaks in $\Theta'_V(\omega)$ and the complex zeros of $V(\omega)$ produce negative peaks in $\Theta'_V(\omega)$. The real poles and zeros of $V(\omega)$ do not significantly affect the peaks in $\Theta'_V(\omega)$. Therefore the contributions of poles and zeros can be separated by considering the positive and negative portions of $\Theta'_V(\omega)$ respectively. Let

$$\Theta'_V(\omega)=[\Theta'_V(\omega)]^{+}+[\Theta'_V(\omega)]^{-} \tag{11}$$

where

$$[\Theta'_V(\omega)]^+ = \begin{cases} \Theta'_V(\omega) & \text{for } \Theta'_V(\omega) \geq 0, \\ 0 & \text{for } \theta'_V(\omega) < 0 \end{cases} \quad (12)$$

and

$$[\Theta'_V(\omega)]^- = \begin{cases} \Theta'_V(\omega) & \text{for } \Theta'_V(\omega) < 0, \\ 0 & \text{for } \Theta'_V(\omega) \geq 0. \end{cases} \quad (13)$$

We can express $[\Theta'_V(\omega)]^+$ and $[\Theta'_V(\omega)]^-$ separately in terms of the cepstral coefficients for poles and zeros as follows: Let

$$[\Theta'_V(\omega)]^+ = C + \sum_{k=1}^{\infty} kc^+(k) \cos k\omega \quad (14)$$

and

$$[\Theta'_V(\omega)]^- = -C + \sum_{k=1}^{\infty} kc^-(k) \cos k\omega, \quad (15)$$

where $\{c^+(k)\}$ and $\{c^-(k)\}$ represent the cepstral coefficients for pole and zero spectra of $V(\omega)$ respectively and C is the average value. Notice that $c(k) = c^+(k) + c^-(k)$, which means that the cepstral coefficients are split into two parts, one corresponding to poles and the other to zeros.

Here $[\Theta'_V(\omega)]^+$ represents the significant portion of the NDPS for the poles of $V(\omega)$ and $[\Theta'_V(\omega)]^-$ represents the significant portion of the NDPS for the zeros of $V(\omega)$. By significant portion we mean that the shape of the curve in the positive portion of $\Theta'_V(\omega)$ is mainly due to poles only and the shape in the negative portion of $\Theta'_V(\omega)$ is meainly due to zeros only. It is very important to note that the shape information is preserved in $c^+(k)$ and $c^-(k)$ for $k = 1, 2, \ldots,$ for poles and zeros respectively.

In most cases of signal analysis, the objective is to represent the smoothed spectrum of a signal by a model. The smoothed spectrum is determined by the first few cepstral coefficients in (8), since they are the first few Fourier coefficients of the log spectrum. If the series are truncated, then the resulting spectrum is called cepstrally smoothed spectrum. It should be noted that the value of $c(0)$ does not affect the shape of the spectrum. Following the same logic, we can obtain cepstrally smoothed spectra for poles and zeros separately by considering only the first few cepstral coefficients in $\{c^+(k)\}$ and $\{c^-(k)\}$ respectively.

In practice there will be some interaction between poles and zeros in the derivative of phase spectrum due to discrete time nature of the signals being considered. The interaction will be more severe of course when the derivative of phase spectrum is computed using only a small number of cepstral coefficients. But still the shapes of the positive and negative portions of the NDPS plot are largely due to contributions of peaks and valleys of the smoothed spectrum respectively. We have found that the interaction between poles and zeros in the NDPS does not significantly affect the resulting model spectrum if a sufficiently large number of cepstral coefficients are considered.

We now describe a method of deriving the parameters of a pole-zero model that represents the smoothed spectrum of a signal. 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, the numerator and the denominator polynomials can be considered as two inverse filters of linear prediction analysis. Consequently $\{a^+(k)\}$ and $\{a^-(k)\}$ represent two sets of linear predictor coefficients (LPCs). The cepstral coefficients of a finite all-pole stable system can be expressed recursively through the LPCs as shown in [6]. The reverse recursion i.e., LPCs from cepstral coefficients is also possible, provided it is known that the cepstral coefficients are for a stable all-pole system. By splitting the cepstral coefficients of the smoothed spectrum into a pole part and a zero part, we achieved a decomposition which enables us to use the reverse recursion to obtain the coefficients of the numerator and denominator polynomials in (1). The pole coefficients $\{a^+(k)\}$ and the zero coefficients $\{a^-(k)\}$ are given by the following relations:

Pole coefficients:

$$a^+(1) = -c^+(1)$$

$$ja^+(j) = -jc^+(j) - \sum_{k=1}^{j-1} kc^+(k)a^+(j-k)$$

$$\text{for } j = 2, 3, \ldots M \quad (16)$$

Zero coefficients:

$$a^-(1)=c^-(1)$$

$$ja^-(j)=jc^-(j)+\sum_{k=1}^{j-1} kc^-(k)a^-(j-k)$$

$$\text{for } j=2,3,\ldots M \tag{17}$$

Only the first M coefficients of $\{c^+(k)\}$ and $\{c^-(k)\}$ are needed to determine completely the parameters of the model given in (16).

3.3. *Error criterion*

Conventionally, the parameters of a pole-zero model are determined using a minimization of mean squared error criterion. Linear prediction analysis has been shown to be equivalent to autocorrelation matching [2]. That is, if $\{R(k)\}$ and $\{\hat{R}(k)\}$ represent the autocorrelation coefficients of a given signal and the impulse response of its all-pole model respectively, then for a pth order model

$$R(k)=\hat{R}(k) \quad \text{for } k=0,1,\ldots p, \tag{18}$$

minimizes the total error E_1 given by

$$E_1=\frac{1}{2\pi}\int_{-\pi}^{\pi}[P(\omega)/\hat{P}(\omega)]\,d\omega \tag{19}$$

where

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

$$\text{(original spectrum)} \tag{20}$$

and

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

$$\text{(model spectrum).} \tag{21}$$

Analogously, if the linear system model is derived from the cepstral coefficients using the relations (16) and (17), then

$$c(k)=\hat{c}(k) \quad \text{for } k=1,2,\ldots p. \tag{22}$$

If the gain term G in $H(z)$ is chosen such that $\ln(G)=\frac{1}{2}c(0)$, then it can be shown that

$$c(0)=\hat{c}(0). \tag{23}$$

The proposed method can thus be interpreted as pole-zero modeling by *cepstral matching*, which can be stated as follows: For a given order M of the pole-zero model, determine the model parameters such that the first $M+1$ cepstral coefficients of the model are equal to the first $M+1$ cepstral coefficients of the signal. The error between the original and the model log spectra is given by

$$E_2=\frac{1}{2\pi}\int_{-\pi}^{\pi}[\ln P(\omega)-\ln\hat{P}(\omega)]^2\,d\omega. \tag{24}$$

Writing E_2 in cepstral coefficients [6], we get

$$E_2=[c(0)-\hat{c}(0)]^2+2\sum_{k=1}^{\infty}[c(k)-\hat{c}(k)]^2. \tag{25}$$

After matching, the error becomes

$$E_2=2\sum_{k=M+1}^{\infty}[c(k)-\hat{c}(k)]^2. \tag{26}$$

It should be noted that there is no minimization process involved in this method. We have only shown that if the cepstral coefficients of the model are chosen so as to match the first $M+1$ cepstral coefficients of the signal, then the resulting mean-squared log spectral error is given by (26). In practice $\hat{c}(k)$ decays as $1/k$ for large k and therefore the value of E_2 is mostly decided by $\{c(k)\}$ alone.

4. Pole-zero decomposition of speech spectra

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

Speech is the putput of a nonstationary vocal tract system, excited either by quasiperiodic glottal pulses or turbulent noise or both. Thus the signal is a convolution of the excitation signal and the impulse response of vocal tract system. Since both the system and the excitation are nonstationary, only short segments (10–40 ms) of speech signal are considered for analysis. During an analysis interval the system and the excitation are assumed to be stationary. The objective in speech analysis is to separate the smoothed spectrum corresponding to the vocal tract system and the fine structure corresponding to the excitation.

In this paper we consider speech signals sampled at 10 kHz. The data is multiplied with a Hamming window before computing the spectrum. The detailed steps of the algorithm for pole-zero decomposition are given in Fig. 2. The derivative of phase spectrum is computed from the first M cepstral coefficients. The choice of M depends on the accuracy of representation required for the spectrum, the accuracy being specified in terms of the number of cepstral coefficients to be matched. The effect of these parameters on the resulting smoothed spectrum is discussed in Section 5. All the DFTs in the algorithm are computed using a 512-point FFT.

5. Results and discussion

In this section we consider several examples of speech spectra to illustrate the application of the proposed method. Our aim here is to show the effectiveness of the method in deriving a pole-zero system that represents the smoothed speech spectrum. Data for these examples was obtained from a

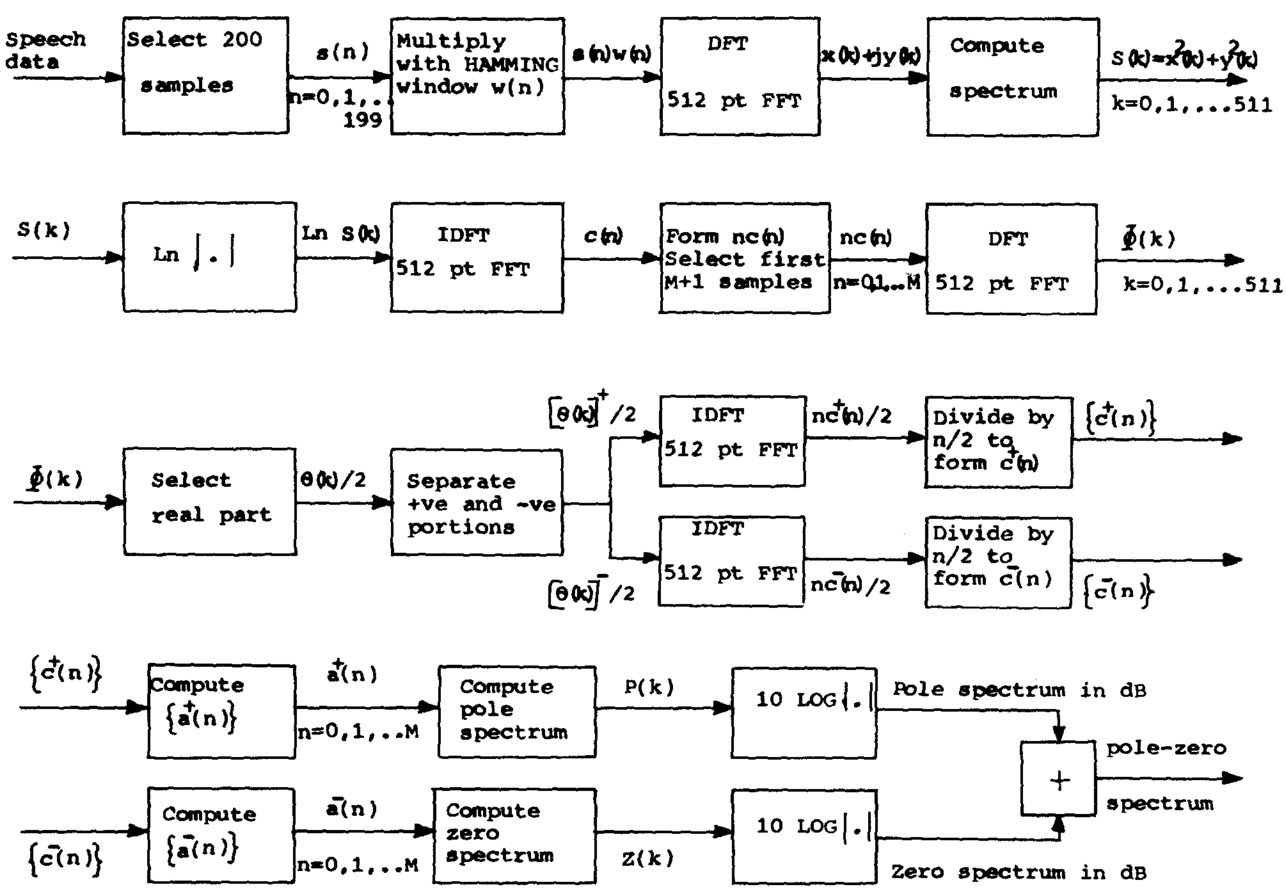

Fig. 2. Block diagram showing computational steps for pole zero decomposition.

spoken utterance, bandpass filtered (80–4500 Hz) and sampled at 10 kHz. A segment of 20 msec (200 samples) was used in the analysis. The sepctrum was computed as described in Section 4.

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. The NDPS for a voiced segment for three different values of $M(10, 20, 30)$ are shown in Fig. 3. The NDPS was obtained by computing the expression

$$\Theta'(\omega) = \sum_{k=1}^{M} kc(k)\cos k\omega. \tag{27}$$

The dotted horizontal line in the figure indicates the dividing line between poles and zeros. The short-time spectrum of the segment is also plotted in the Fig. 3. It can be observed that positive peaks in the NDPS plot correspond to peaks in the smoothed short-time spectrum. Similarly negative peaks in the NDPS correspond to dips in the smoothed spectrum. The improvement in resolution for higher values of M is also evident from Fig. 3. The pole spectrum $P(\omega)$, the zero spectrum $Z(\omega)$ and the pole-zero spectrum $P(\omega)Z(\omega)$ for $M = 20$ are shown in Fig. 4. The various log spectra in dB are computed as follows:

Pole spectrum:

$$10\log P(\omega) = 10\log\left[1\Big/\left|1+\sum_{k=1}^{M} a^{+}(k)\,e^{-j\omega k}\right|^{2}\right] \tag{28}$$

Zero spectrum:

$$10\log Z(\omega) = 10\log\left[\left|1+\sum_{k=1}^{M} a^{-}(k)\,e^{-j\omega k}\right|^{2}\right] \tag{29}$$

Pole-zero spectrum:

$$10\log|H(\omega)|^2 = 10\log P(\omega) + 10\log Z(\omega). \tag{30}$$

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. In this way the pole-zero spectrum provides a uniformly accurate representation of the overall smoothed spectrum. In Fig. 5 the pole-zero spectrum is superimposed on the short time spectrum

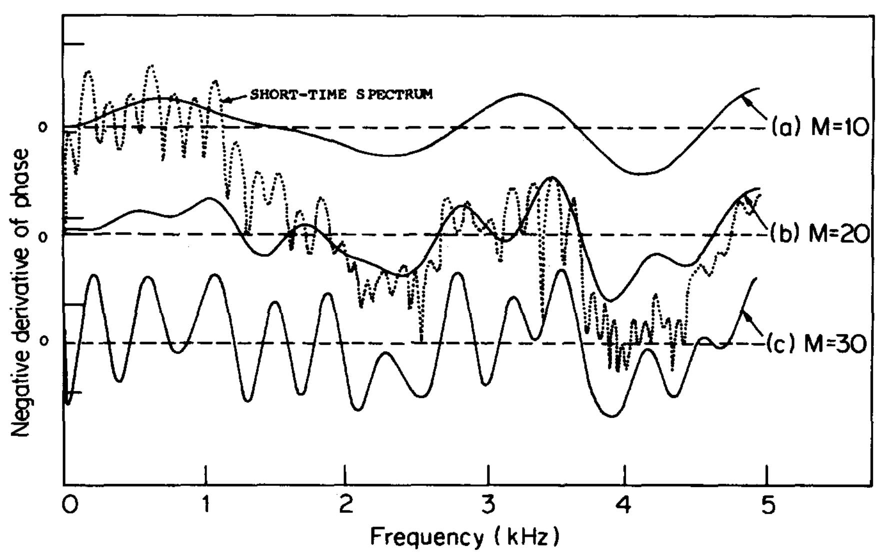

Fig. 3. Negative derivative of phase spectra for a segment of voiced speech for different values of M. The short-time spectrum of the segment is also shown in the figure by a dotted curve. (a) $M = 10$, (b) $M = 20$, (c) $M = 30$.

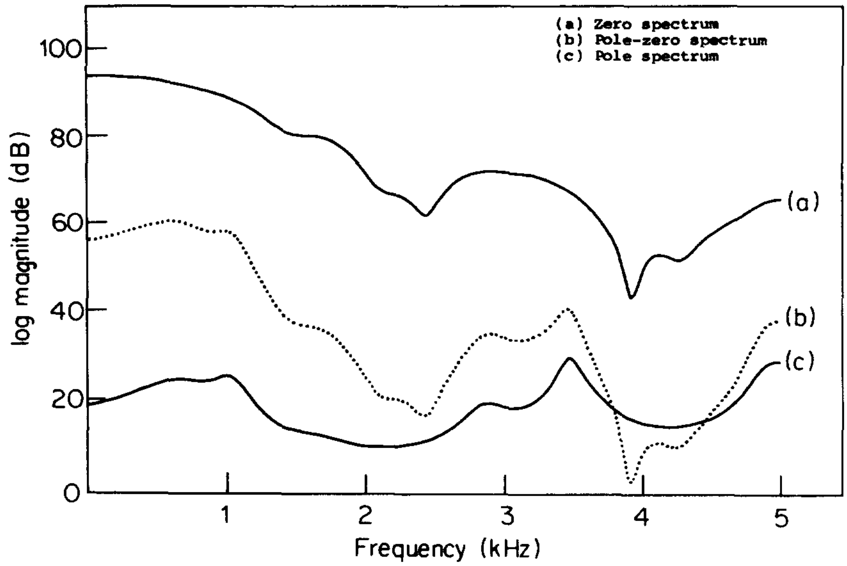

Fig. 4. Component spectra of a pole-zero model ($M = 20$) for a segment of voiced speech. The pole-zero spectrum is obtained by adding the pole and zero spectra on dB scale. (a) Zero spectrum, (b) pole-zero spectrum, (c) pole spectrum.

of speech to illustrate the nature of fit of the model spectrum.

The value of M determines the resolution in the smoothed spectrum. Fig. 6 shows the pole-zero spectra for different values of M. Although some improvement in resolution is noticed for higher values of M, the shape of model spectrum remains essentially the same for all values of M.

Comparison of our pole-zero analysis with linear prediction analysis is made in the derivative of phase spectral domain. For this purpose the NDPS for the original data, pole-zero model and

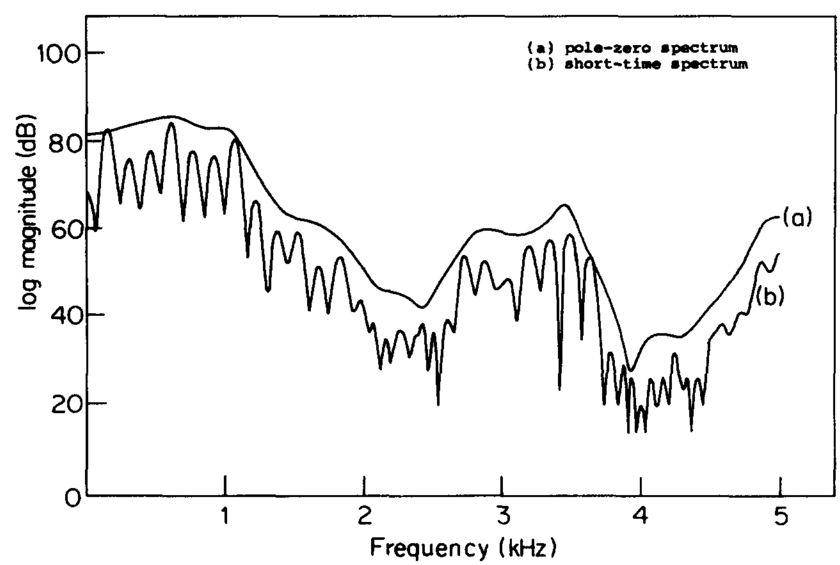

Fig. 5. Pole-zero spectrum for $M = 20$ superimposed on the short-time spectrum of the voiced segment of speech.

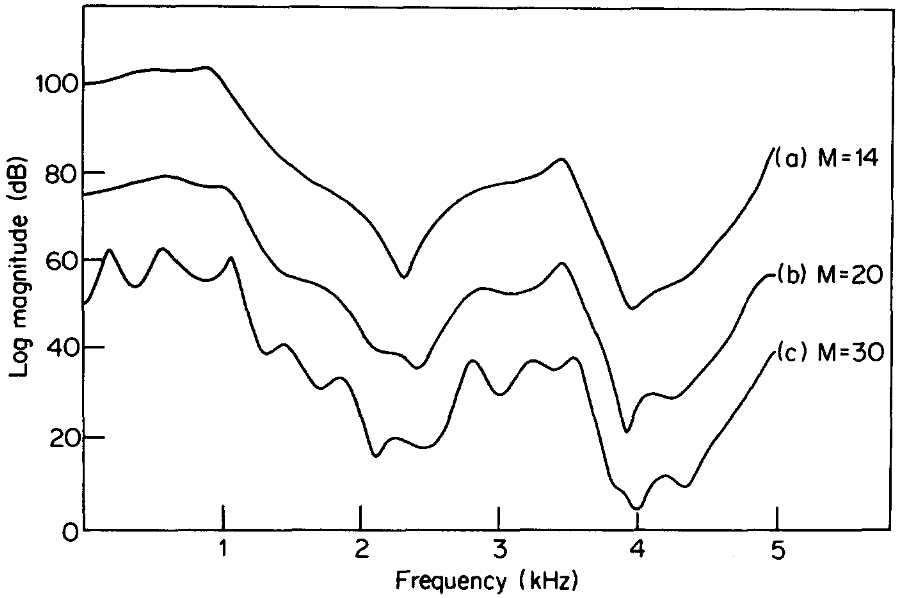

Fig. 6. Pole-zero model spectra for different values of M for a segment of voiced speech. (a) $M = 14$, (b) $M = 20$, (c) $M = 30$.

all pole model are plotted in Fig. 7. To maintain the same resolution, a 20th order all pole filter is compared with a 20th order (20 poles and 20 zeros) pole-zero model. The NDPS of the original spectrum is approximated well by the pole-zero model because the first M Fourier coefficients of the two plots are exactly equal. The sharper peaks and valleys in the NDPS plot for the pole-zero model are due to extrapolation of the cepstral coefficients beyond $M = 20$. On the other hand,

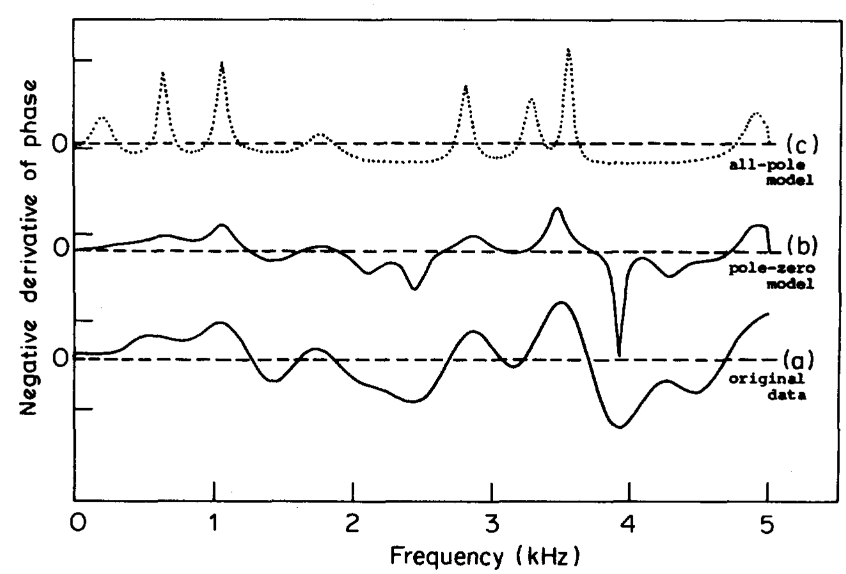

Fig. 7. Comparison of pole-zero model and all pole model in the NDPS domain. (a) Original data, (b) pole-zero model ($M = 20$), (c) all pole model ($M = 20$).

only peaks are significant in the NDPS plot for the all pole filter and the peaks are much sharper than for the pole-zero model.

Results of analysis for a segment of voiced fricative are shown in Fig. 8. A 20-pole spectrum obtained by linear prediction analysis is also shown along with the pole-zero spectrum for comparison. The first spectral peak in the all pole model spectrum corresponds to the fundamental frequency. Such a peak is absent in the pole-zero model spectrum. In general, a peak in the smoothed spectrum can also occur due to two closely spaced zeros, and hence cannot always be considered that all peaks correspond to resonances only. This point is illustrated in the figure where the closely spaced zeros near 2.3 kHz and 2.7 kHz produced a sharp peak at 2.5 kHz. The approximation of the zeros by the pole-zero model is clearly demonstrated.

Spectral fit improves as the order of M is increased, but as M is made very large, the original spectrum inclusive of the fine structure due to source also appears. Since the cepstral coefficients for large quefrencies have negligible components due to vocal tract system, by considering only the high quefrency portion of the cepstrum, the excitation information can be obtained. The NDPS for this purpose is computed using the formula

$$\Theta'(\omega) = \sum_{k=21}^{255} kc(k) \cos k\omega. \tag{33}$$

The plot of $\Theta'(\omega)$ for a vowel segment is given in Fig. 9. The figure illustrates the ability of the derivative of phase spectrum in resolving even the fine structure of the spectrum. We are currently exploring the possibility of using this property for reliable pitch estimation [7].

6. Conclusions

A new technique for representing the smoothed spectrum of speech by pole-zero models has been presented. For each specified match in the cepstral domain a pole-zero model is obtained in a straight forward manner. The elegance of the method lies in determining the parameters of the model uniquely by matching a specified number of cepstral coefficients of a given signal. This technique can be called "cepstral matching" analogous to autocorrelation matching in all pole modeling [2].

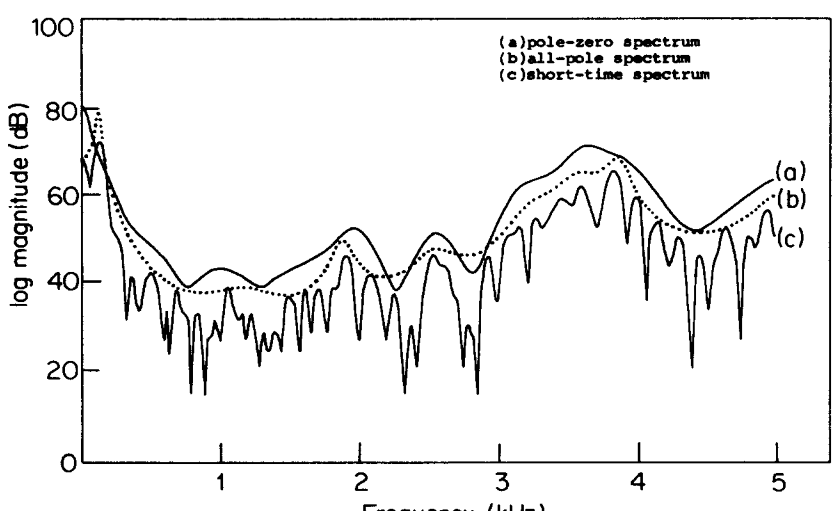

Fig. 8. Comparison of pole-zero spectrum with the all pole spectrum of LP analysis for a segment of voiced fricative. The short-time spectrum of the segment is also shown in the figure. (a) Pole-zero spectrum ($M = 20$), (b) all-pole spectrum ($M = 20$), (c) short-time spectrum.

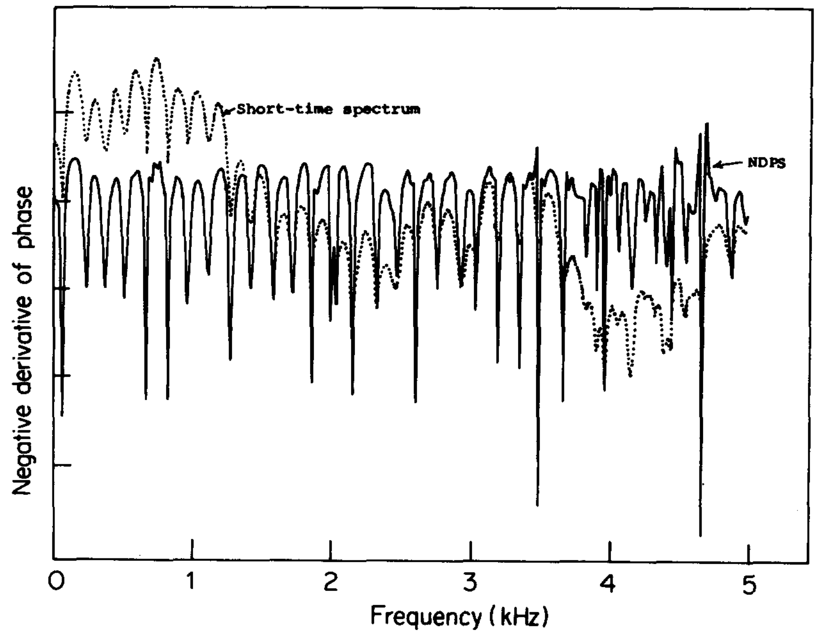

Fig. 9. NDPS for a segment of voiced speech computed from equation (33). The short-time spectrum of the segment is also shown in the figure to demonstrate the resolution of spectral peaks and valleys in the NDPS.

The method of obtaining the smoothed spectrum is different from the conventional cepstral smoothing [3] where the cepstral coefficients are truncated. In the present method the cepstral coefficients beyond the specified order are uniquely extrapolated to improve the frequency resolution.

For purposes of speech compression it is possible to represent each of the component spectra i.e., the pole spectrum and the zero spectrum by a smaller number of coefficients than M poles and M zeros. Since the pole spectrum will usually have 3 to 5 significant peaks, it is possible to represent the spectrum with 8 to 12 coefficients. These coefficients can be derived from all pole modeling of the pole spectrum. Similarly, the inverse of the zero spectrum will usually have 2 to 3 significant peaks, and it is possible to represent it with 6 to 8 coefficients using all pole modeling. Thus the short time spectrum can be represented effectively by about 14 to 20 coefficients.

An extremely useful property of the pole part of the derivative of phase spectrum is that it is nonzero in the frequency regions corresponding to peaks of the smoothed spectrum. Normally, regions around peaks in the smoothed spectrum can be used to represent high signal to noise portions of the spectrum. The NDPS representation of the smoothed spectrum provides a method of determining such regions automatically.

There appears to be good potential in the approach for solving a variety of problems encountered in the field of digital signal processing, like for example, the design of digital filters, decomposition of composite signals, deconvolution of convolved signals. Presently some of these applications are being studied.

Acknowledgments

The author wishes to thank Prof. D. Raj Reddy, Dr. T.V. Ananthapadmanabha and Dr. R.G. Goodman for many helpful comments.

This research was sponsored by the Defence Advanced Research Projects Agency (DOD), ARPA Order No. 3597, and monitored by the Air

Force Avionics Laboratory under contract F33615-78-C1151.

The views and the conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Defence Advanced Research Projects Agency or the U.S. Government.

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, 1978, pp. 1310–1318.

[2] J.I. Makhoul, "Linear prediction: A Tutorial Review," *Proc. IEEE*, Vol. 63, April 1975, pp. 561–580.

[3] A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975, ch. 10.

[4] B. Yegnanarayana, "Formant extraction from linear prediction phase spectra," *J. Acoust. Soc. Amer.*, Vol. 63, 1978, pp. 1638–1640.

[5] A.J. Berkhout, "Related properties of minimum-phase and zero-phase time functions," *Geophysical Prospecting*, Vol. XXII, No. 4, 1974, pp. 683–709.

[6] A.H. Gray and J.D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust. Speech and Signal Processing*, Vol. ASSP-24, October 1976, pp. 380–391.

[7] B. Yegnanarayana and T.V. Ananthapadmanabha, "On improving the reliability of cepstral pitch extraction," Dept. of Computer Science, Carnegie-Mellon University, Pittsburgh, PA 15213, Report CMU-CS-79-108, February 1979.