

NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:
The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

CORRECTIONS TO
 "Pole-zero Decomposition of Speech Spectra"
 B. Yegnanarayana

ON PAGE 5

1. In Sec.3.1 the following sentence should be added after the first sentence.
Throughout this paper we consider only discrete signals (i.e., signals sampled at Nyquist rate), so that the Fourier transform is periodic in ω with period 2π .
2. In Sec.3.1 $\log V(\omega)$ should be replaced by $\ln V(\omega)$.

ON PAGE 6

1. In equation (8) \log should be replaced by \ln .
2. The last line should read as follows:
poles and zeros, since the relation in (10) is valid for poles or zeros or for both, except for a constant term (C) on the right hand side. Let

ON PAGE 7

1. In equation (14) a constant term $-C$ should be added to the right hand side.
2. In equation (15) a constant term C should be added to the right hand side.

ON PAGE 8

Equation (16) should be changed as follows:

$$a^{-1} = c^{-1}$$

$$j a^{-j} = j c^{-j} + \sum_{n=1}^{j-1} n c^{-n} a^{-j-n}, \quad \text{for } j=2,3,\dots,M_p \quad (16)$$

ON PAGE 12

The fourth line from the bottom should read as follows:
for different values of M. This may be due to the fact that the choice of M_p equal to 20 has

ON PAGE 17

1. In equation (A2) \log should be replaced by \ln .
2. The right hand side of equation (A3) should be multiplied by -1 .

FIG.2

In Fig.2 $\text{LOG } S(k)$ should be replaced by $\text{Ln } S(k)$.

CMU-CS-79-101

POLE-ZERO DECOMPOSITION OF SPEECH SPECTRA

B.Yegnanarayana
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213.

JANUARY 1979

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.

ABSTRACT

A new method for determining the parameters of a pole-zero model for speech spectra is proposed in this paper. In this method the cepstral coefficients of a signal are split into two parts, one corresponding to poles and the other to zeros. The decomposition is achieved by using the properties of the derivative of phase spectra of minimum phase signals. Parameters of the model are derived recursively from the cepstral coefficients for poles and zeros separately. Since poles and zeros are treated alike and derived independently, there is no effect of one on the other. The method is illustrated with several examples of speech spectra. It is shown that in all cases the envelope fit is equally good at peaks as well as at valleys in the spectrum. Results of this paper suggest a method of obtaining a linear system model for a given signal using a criterion different from the conventional minimization of mean squared error criterion. Although the method is described for minimum phase signals only, extension of the method to mixed phase signals is trivial, since a mixed phase signal can be split into minimum and maximum phase components using complex cepstrum.

I. INTRODUCTION

An important problem in signal analysis is the estimation of parameters of a pole-zero model for a given signal spectrum. In this paper we present a general and effective method for determining the parameters. The method involves separating the effects of poles and zeros based on the properties of the derivative of linear prediction phase spectra reported recently by the author [1]. Besides the effectiveness of the derivative of phase spectrum, the inherent advantages of linear prediction (LP) and homomorphic filtering approaches are exploited to derive the parameters of the model in a simple and elegant manner. The proposed technique yields a linear system model for a signal using a criterion different from the conventional minimization of mean squared error criterion. Although the technique can be applied to a general class of signals, we confine our discussion to examples from speech signal analysis, as these examples provide the necessary diversity of situations and also have a physical interpretation.

Approximating speech spectra by pole-zero models and estimating the parameters of such models has recently been the subject of active research [2]-[6]. Techniques for solving this estimation problem are primarily based on the principles of linear prediction and homomorphic deconvolution. The approach in these methods is to determine pole parameters first, either directly from the signal or from the minimum phase equivalent of the signal. Zero parameters are then determined from the residual signal by one of the several well known methods [7],[9]. Since pole-zero modelling is a classical problem in the general area of system identification, extensive discussion of this problem can also be found in linear systems literature. Methods for simultaneously estimating pole-zero parameters are generally iterative in nature and little is known theoretically about the convergence properties of the algorithms available [3],[10],[11].

In this paper a simple method for simultaneously determining the poles and zeros of a model is presented. This method considers poles and zeros in an identical manner. In Sec.II the problem of pole-zero estimation and the underlying principle of the proposed technique

are discussed. In Sec.III the technique for complete pole-zero decomposition is presented. An algorithm for pole-zero decomposition of speech spectra is presented in Sec.IV. Several 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. Some issues presently under investigation are cited in Sec.VI.

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.

2.1 The Problem

For a given signal $x(n)$, determine the parameters of the pole-zero model

$$H(z) = N(z)/D(z) \quad (1)$$

where

$$N(z) = \sum_{n=0}^{M_z} a^+(n) z^{-n} \quad (2)$$

and

$$D(z) = \sum_{n=0}^{M_p} a^-(n) z^{-n}, \quad (3)$$

such that the frequency response of the model matches the *envelope* of the spectrum of $x(n)$.

In earlier attempts to solve this problem, a minimization criterion is invariably used. Of particular interest for problems arising in speech are the linear prediction analysis and the homomorphic filtering. In linear prediction analysis, the best all-pole filter (i.e., $H(z)$ when $N(z)=1$) is obtained by using a minimum mean squared error criterion. The error is computed as the difference between signal samples and their linearly predicted values. In the spectral domain this criterion is equivalent to minimization of the integrated ratio of the signal

spectrum and the model spectrum [7]. In homomorphic filtering, a minimum phase estimate $v(n)$ of $x(n)$ is initially obtained [8]. This avoids the problem of pitch synchronization for analysis of speech signals, since $v(n)$ is a minimum phase estimate of the vocal tract impulse response [4]. Linear prediction analysis is then performed on $v(n)$ to determine the poles of the model. This method is called homomorphic prediction [5]. Zeros of the model are then determined by repeated application of linear prediction analysis [2] or by Shank's method [9] or by inverse linear prediction [7]. An obvious limitation in these methods is that the accuracy of the estimated poles is affected by the presence of zeros in the signal spectrum. Methods for simultaneous estimation of poles and zeros are iterative [11],[12] and computationally complex. An interesting method for separating poles and zeros is by cepstral prediction [5]. This method uses the complex cepstrum $\hat{x}(n)$ of $x(n)$. It can be shown that the poles of $n\hat{x}(n)$ correspond to the poles and zeros of the original signal $x(n)$. Linear prediction analysis is performed to determine all the poles of $n\hat{x}(n)$. Each pole of $n\hat{x}(n)$ is classified as either a pole or a zero of $x(n)$ by observing the sign of the residue of the z-transform of $n\hat{x}(n)$ at the pole [5]. However, this method also involves estimation of poles from a signal having both poles and zeros.

2.2 Basis for Pole-zero Decomposition

Since the objective in the present problem is to determine a pole-zero model which fits a *spectral envelope*, 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 [8],[13]. In particular, all poles and zeros of a minimum phase signal lie within the unit circle in the z-plane.

Properties of the *derivative* of phase spectrum of a stable all-pole system have been recently reported by the author[1]. These properties were suggested for extraction of formants 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. The derivative of phase spectrum of a typical first order filter

(real pole) is given by

$$\theta_1'(\omega) = -\gamma/(\omega^2 + \gamma^2) \quad (4)$$

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

$$\theta_2'(\omega) = -2\alpha(\alpha^2 + \beta^2 + \omega^2) / \{(\alpha^2 + \beta^2 - \omega^2)^2 + 4\omega^2\alpha^2\} \quad (5)$$

where α and β are the half power bandwidth and resonance frequency of the filter. These equations are derived in [1]. In general $\beta^2 \gg \alpha^2$. The derivative of phase spectrum 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 given below.

1. $-\theta_1'(\omega)$ is a monotonically decreasing function of ω
2. At low frequencies $\theta_1'(\omega) \approx -1/\gamma$
3. At high frequencies $\theta_1'(\omega) \approx -\gamma/\omega^2$
4. $-\theta_2'(\omega)$ is approximately proportional to the squared magnitude response of the filter around the resonance frequency
5. At low frequencies $\theta_2'(\omega) \approx -2\alpha/\beta^2$, which is a small constant quantity
6. At high frequencies $\theta_2'(\omega) \approx -2\alpha/\omega^2$

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 similar behaviour for real and complex conjugate zeros in the derivative of phase spectra. The only difference is that the derivative of phase response for zeros will have a sign opposite to that for poles. Specifically, $\theta'(\omega)$ will have a negative peak

due to a complex conjugate pole pair and a positive 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 negative derivative of phase spectra for a first order 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 poles and to frequencies near the resonance frequency for complex conjugate poles. In this paper all plots of the derivative of phase spectra are shown after multiplying with -1 , so that positive peaks in the plots can be compared with peaks in the magnitude spectrum, which correspond to resonances.

III. 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. Since all poles and zeros of $V(\omega)$ lie within the unit circle in the z -plane [8], $\log V(\omega)$ can be expressed in Fourier series expansion as follows:

$$\log V(\omega) = c(0) + \sum_{n=1}^{\infty} c(n) e^{-j\omega n} \quad (6)$$

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

$$V(\omega) = |V(\omega)| e^{j\theta_V(\omega)}, \quad (7)$$

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

$$\log |V(\omega)| = \sum_{n=0}^{\infty} c(n) \cos n\omega \quad (\text{real part}) \quad (8)$$

and

$$\theta_V(\omega) + 2\lambda\pi = - \sum_{n=1}^{\infty} c(n) \sin n\omega \quad (\text{imaginary part}) \quad (9)$$

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

$$\theta'_V(\omega) = - \sum_{n=1}^{\infty} nc(n) \cos n\omega. \quad (10)$$

3.2 Pole-zero Decomposition

$\theta'_V(\omega)$ is the derivative of phase spectrum of a minimum phase signal whose properties were discussed in Sec.II. In particular, the complex conjugate poles of $V(\omega)$ produce negative peaks in $\theta'_V(\omega)$ and the complex conjugate zeros of $V(\omega)$ produce positive 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 negative and positive portions of $\theta'_V(\omega)$ respectively. Let

$$\theta'_V(\omega) = [\theta'_V(\omega)]^- + [\theta'_V(\omega)]^+ \quad (11)$$

where

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

and

$$\begin{aligned} [\theta'_V(\omega)]^+ &= \theta'_V(\omega) & \text{for } \theta'_V(\omega) \geq 0 \\ &= 0 & \text{for } \theta'_V(\omega) < 0. \end{aligned} \quad (13)$$

We can express $[\theta'_V(\omega)]^-$ and $[\theta'_V(\omega)]^+$ separately in terms of the cepstral coefficients for poles and zeros, since the relation given in (10) is valid for poles or zeros or for both. Let

$$[\theta'_V(\omega)]^- = - \sum_{n=1}^{\infty} n c^-(n) \cos n\omega \quad (14)$$

and

$$[\theta'_V(\omega)]^+ = - \sum_{n=1}^{\infty} n c^+(n) \cos n\omega, \quad (15)$$

where $\{c^-(n)\}$ and $\{c^+(n)\}$ represent the cepstral coefficients for pole and zero spectra of $V(\omega)$ respectively. Notice that $c(n) = c^-(n) + c^+(n)$, 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 derivative of phase spectrum for the poles of $V(\omega)$ and $[\theta'_V(\omega)]^+$ represents the significant portion of the derivative of phase spectrum 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 largely due to zeros only and the negative portion of $\theta'_V(\omega)$ is largely due to poles only. It is very important, for later discussion, to note that the shape information is preserved in $c^-(n)$ and $c^+(n)$ for $n=1, 2, \dots$, for poles and zeros respectively.

In most cases of signal analysis, the objective is to determine the *envelope* of a signal spectrum. The spectral envelope 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 the 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 the cepstrally smoothed spectra for poles and zeros separately by considering only the first few cepstral coefficients in $\{c^-(n)\}$ and $\{c^+(n)\}$ respectively.

We now describe a method of deriving the parameters of a pole-zero model that represents the envelope of a signal spectrum. 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 for a minimum phase spectral envelope, the numerator and the denominator polynomials can be considered as two inverse filters of linear prediction analysis [14].

Consequently, $\{a^-(n)\}$ and $\{a^+(n)\}$ represent two sets of LPCs. The cepstral coefficients of a finite all-pole stable system can be expressed recursively through the LPCs as shown in [15]. These relations are also given in the Appendix. The inverse 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 envelope spectrum into a pole part and a zero part, we achieved a decomposition which enables us to use the inverse recursion to obtain the coefficients of the numerator and denominator polynomials in (1). The pole coefficients $\{a^-(n)\}$ and the zero coefficients $\{a^+(n)\}$ are given by the following relations:

Pole Coefficients:

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

$$j a^-(j) = -j c^-(j) - \sum_{n=1}^{j-1} c^-(n) a^-(j-n), \quad \text{for } j=2,3,\dots,M_p \quad (16)$$

Zero Coefficients:

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

$$j a^+(j) = -j c^+(j) - \sum_{n=1}^{j-1} c^+(n) a^+(j-n), \quad \text{for } j=2,3,\dots,M_z \quad (17)$$

Only M_p coefficients of $\{c^-(n)\}$ and M_z coefficients of $\{c^+(n)\}$ are needed to determine completely the parameters of the model given in (16). The choice of M_p and M_z for speech signals is discussed in Sec.V.

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 [6]. That is, if $\{R(n)\}$ and $\{\hat{R}(n)\}$ represent the autocorrelation coefficients of a given signal and the impulse response of its all-pole model respectively, then for a p-th order model,

$$R(n) = \hat{R}(n) \quad \text{for } n=0,1,\dots,p, \quad (18)$$

minimizes the total error E_1 given by

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

where

$$P(\omega) = \sum_{n=0}^{\infty} R(n) \cos n\omega \quad (\text{original spectrum}) \quad (20)$$

and

$$\hat{P}(\omega) = \sum_{n=0}^{\infty} \hat{R}(n) \cos n\omega \quad (\text{model spectrum}). \quad (21)$$

Analogously, if the linear system model is derived from the cepstral coefficients using the relations given in the Appendix, then

$$c(n) = \hat{c}(n) \quad \text{for } n=1,2,\dots,p. \quad (22)$$

If the energy in the original spectrum and the model spectrum are equal, then

$$c(0) = \hat{c}(0) \quad (23)$$

because $c(0) = \log R(0)$ and $\hat{c}(0) = \log \hat{R}(0)$.

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

$$E_2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log P(\omega) - \log \hat{P}(\omega)]^2 d\omega \quad (24)$$

Writing E_2 in cepstral coefficients [15], we get

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

After matching, the error becomes

$$E_2 = \sum_{n=M_p+1}^{\infty} [c(n) - \hat{c}(n)]^2. \quad (26)$$

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_p+1 cepstral coefficients of the signal, then the resulting rms log spectral error is given by (26).

IV. IMPLEMENTATION OF POLE-ZERO DECOMPOSITION FOR 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 output 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 excitation are nonstationary, only short segments (10-40ms) of speech signal are considered for analysis. During an analysis interval the system and excitation are assumed to be stationary. The objective in speech analysis is to separate the spectral envelope corresponding to vocal tract system, and the fine structure corresponding to excitation. In most applications of speech analysis it is sufficient if the spectral characteristics are represented accurately. Hence a minimum phase version of the signal is adequate for pole-zero modelling.

In this paper we consider speech signals sampled at 10 kHz. The data was passed through a preemphasis filter $(1 - 0.92z^{-1})$ and then 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 only from M cepstral coefficients. The choice of M , M_p and M_z 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 envelope is discussed in Sec.V. All the DFTs in the algorithm were computed using a 512-point FFT.

V. 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 *envelope* of speech spectrum. The choice of the parameters M , M_p , and M_z and their effect on the resulting envelope are discussed. Data for these examples was obtained from a 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 spectrum was computed as described in Sec.IV.

5.1 Choice of M , M_p , M_z

These parameters determine the resolution of peaks and valleys in the spectral envelope and also the error between the actual spectrum and the modelled spectrum. 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 spectrum with increased resolution for peaks and valleys in the derivative of phase spectrum. The derivative of phase spectrum for a voiced segment for three different values of M (10,20,30) are shown in Fig.3. The derivative of phase spectrum was obtained by computing the expression

$$\theta'(\omega) = - \sum_{n=1}^M nc(n) \cos n\omega. \quad (27)$$

In Fig.3 $-\theta'(\omega)$ are plotted so that the positive part corresponds to poles and the negative part to zeros. The dotted horizontal line indicates the dividing line between poles and zeros and in this case happens to be the x-axis itself. The short-time spectrum of the segment is also plotted in the Fig.3. It can be observed that positive peaks in the derivative of phase spectrum plot correspond to peaks in the envelope of the short-time spectrum. Similarly negative peaks in the derivative of phase spectrum correspond to dips in the envelope. 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 combined envelope spectrum $P(\omega)Z(\omega)$ for different values of M are shown in Figs.4-6. The various log spectra in dB are computed as follows:

Pole spectrum:

$$10 \log P(\omega) = 10 \log [1/|1 + \sum_{n=1}^{M_p} a^-(n) e^{j\omega n}|^2] \quad (28)$$

Zero spectrum:

$$10 \log Z(\omega) = 10 \log [1 + \sum_{n=1}^{M_z} a^+(n) e^{j\omega n}|^2] \quad (29)$$

Pole-zero spectrum:

$$10 \log P(\omega) = 10 \log P(\omega) + 10 \log Z(\omega) \quad (30)$$

For all the cases in Figs.4-6 $M_p=M_z=20$. The resulting spectral envelope is nearly same for different values of M. This may be due to the fact that the choice of M equal to 20 has resolved all the significant spectral peaks. An important design consideration is demonstrated by the analysis of a voiced fricative segment. Spectral envelopes obtained for different values of $M_p(=M_z)$ when $M=40$ are shown in Fig.7. For low values of M_p and M_z , spurious

peaks which do not match with the spectral envelope occur. These peaks have occurred because after using a large value for M , if smaller values of M_p and M_z are used, it is equivalent to truncating the coefficients of an all-zero filter. That is, if the actual filter is of the order M_0 and is given by

$$A(z) = \sum_{n=0}^{M_0} a(n) z^{-n}, \quad (31)$$

then the response that will be computed for low values of M_p is for the filter

$$A(z) = \sum_{n=0}^{M_p} a(n) z^{-n}. \quad (32)$$

In the above expression, the last $M_0 - M_p$ values of $\{a(k)\}$ are removed by truncation. Such an effect will not be observed if M_p is sufficiently large. This result may also provide a clue to determine the actual number of poles and zeros present for a given spectral envelope. The truncation effect will disappear if $M = M_p = M_z$. This is demonstrated in Fig.8 where the envelope spectra are plotted for the three values of M_p considered in Fig.7, but with the restriction that $M = M_p = M_z$. Now even for low values of M_p there are no spurious peaks as in Fig.7. The resulting spectra in Fig.8 should be interpreted as smoothed versions for the specified order of the model. Thus, low values of $M_p (=M_z)$ for a given large value of M result in truncation, whereas low values of M_p subjected to the condition $M = M_p = M_z$ result in smoothing the spectrum appropriately.

5.2 Examples

Results of analysis for four segments of speech sounds corresponding to a diphthong, a nasal, a voiced fricative and an unvoiced fricative are shown in Figs.9-12. In all these cases $M=40$ and $M_p=M_z=20$. In these figures the short-time spectrum is shown by thin solid curve and the response of the pole-zero model by thick solid curve. The smoothed spectrum obtained by LP analysis is also shown by dotted curve for comparison. A 20-th order predictor was used. The effectiveness with which the dips in the spectrum are represented by the present method can be noted from Fig.9. The spectral dips around 3 kHz and 4 kHz

are represented better in the pole-zero spectrum than in the LP spectrum. LP method introduces some spurious peaks in the spectrum which do not match with the envelope of the short-time spectrum. Of course as the order of the pole-zero model is creased, additional peaks will appear in its response, approximating the short-time spectrum in more detail.

Result of analysis for a nasal segment is shown in Fig.10. The first spectral peak in the pole-zero spectrum is much broader compared to the peak in the LP spectrum. In fact the first spectral peak in the LP spectrum corresponds to the fundamental frequency. Further, the zero at the origin sharpens the supposedly resonant peak in the LP spectrum. In general a peak in the spectral envelope can also occur due to two closely spaced zeros, and hence it cannot always be considered that all peaks correspond to resonances only. This point is illustrated in Fig.11, where the pole-zero spectrum for a voiced fricative is plotted. The two closely spaced zeros near 2.3 kHz and 2.7 kHz produced a sharp peak at 2.5 kHz. Very good representation of valleys in the spectral envelope is obtained even for the case of an unvoiced fricative as shown in Fig.12.

In general spectral fit improves as the order of M is incresed, 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 quefrenicy have negligible components due to vocal tract system, by considering only the high quefrenicy portion of the cepstrum, the excitation information can be obtained. The derivative of phase spectrum for this purpose is computed using the formula

$$\theta'(\omega) = - \sum_{n=21}^{255} n c(n) \cos n\omega. \quad (33)$$

The positive and negative parts of $-\theta'(\omega)$ for a voiced segment are plotted separately in Fig.13. 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.

VI. CONCLUSIONS

A new technique for determining the envelope of speech spectrum by a pole-zero model has been presented. For a specified match in the cepstral domain, the pole-zero model is obtained in a deterministic manner. This method not only affords a different approach to speech analysis, but it also provides an *exact* solution to the difficult problem of pole-zero modelling of speech spectrum. The elegance of the method lies in determining the parameters of a model easily while matching a specified number of cepstral coefficients of the model with the cepstral coefficients of a given signal. This technique can be called *cepstral matching* analogous to autocorrelation matching in all-pole modelling [6]. The envelope fit obtained by the present method is different from conventional cepstral smoothing [16], 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.

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 envelope spectrum. Normally regions around peaks in the envelope spectrum are used to represent high signal to noise portions of the spectrum. Therefore now we have an automatic method of determining such regions from the derivative of phase spectrum.

Although modelling for minimum phase signals only is considered here, extension of the method to mixed phase signals is trivial. This is because a mixed phase signal can be split into minimum phase and maximum phase components using complex cepstrum [5] and a

pole-zero model for each component can be determined in an identical manner. It should be noted, however, that for a maximum phase signal the poles and zeros lie within the unit circle in the z^{-1} -plane. There appears to be good potential in the approach for solving a variety of problems commonly 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.

APPENDIX

Let $A(z)$ be the digital inverse filter for an all-pole model given by

$$A(z) = \sum_{k=0}^p a(k) z^{-k} \quad (A1)$$

with $a(0)=1$. If $\{c(n)\}$ represent the cepstral coefficients for $\{a(n)\}$, then by definition it follows

$$\log \{ A(z) \} = - \sum_{k=1}^{\infty} c(k) z^{-k} \quad (A2)$$

Differentiating both sides of (A2) with respect to z , and rearranging the terms, we get

$$\sum_{n=1}^p n a(n) z^{-n} = \left[\sum_{n=1}^{\infty} n c(n) z^{-n} \right] \left[\sum_{n=0}^p a(n) z^{-n} \right] \quad (A3)$$

Multiplying out the series and using the fact that $a(0)=1$, we get

$$\begin{aligned} a(1) &= -c(1), \\ j a(j) &= -j c(j) - \sum_{n=1}^{j-1} n c(n) a(j-n), \quad j=2,3,\dots,p. \end{aligned} \quad (A4)$$

It is to be noted that using (A4) we can obtain $\{a(n)\}$ from $\{c(n)\}$ or vice versa. Only p values of $\{c(n)\}$ are required to determine $\{a(n)\}$. However if $\{a(n)\}$ is known, then all the cepstral coefficients can be determined. Let us denote these values as $\{c'(n)\}$. The first p values of $\{c'(n)\}$ are obtained using the equations:

$$\begin{aligned} c'(1) &= -a(1), \\ j c'(j) &= -j a(j) - \sum_{n=1}^{j-1} n c'(n) a(j-n), \quad j=2,3,\dots,p. \end{aligned} \quad (A5)$$

The remaining values of $\{c'(n)\}$ are obtained by using the relations

$$j c'(j) = - \sum_{n=1}^p (j-n) c'(j-n) a(n), \quad \text{for } j=p+1, p+2, \dots \quad (\text{A6})$$

which is derived from (A3). It is obvious that

$$c(n) = c'(n), \quad \text{for } n=1, 2, \dots, p \quad (\text{A7})$$

and

$$c(n) \neq c'(n), \quad \text{for } n=p+1, p+2, \dots \quad (\text{A8})$$

ACKNOWLEDGMENTS

The author wishes to thank Prof. D. Raj Reddy, Dr. T. V. Ananthapadmanabha and Dr. R.G. Goodman for many helpful comments. In particular, Dr. Ananthapadmanabha's constructive criticism has made it possible to improve the presentation of this work considerably.

REFERENCES

1. B. Yegnanarayana, "Formant extraction from linear prediction phase spectra," *J. Acoust. Soc. Amer.*, vol.63, pp.1638-1640, 1978.
2. 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.
3. K. Steiglitz, "On the simultaneous estimation of poles and zeros in speech analysis," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-25, pp.229-234, June 1977.
4. G.E. Kopec, A.V. Oppenheim and J.M. Tribolet, "Speech analysis by homomorphic prediction," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-25, pp. 40-49, Feb. 1977.
5. A.V. Oppenheim, G.E. Kopec and J.M. Tribolet, "Signal analysis by homomorphic prediction," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-24, pp. 327-332, Aug. 1976.
6. J.I. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp.561-580, April 1975.
7. J.I. Makhoul, "Selective linear prediction and analysis-by-synthesis in speech analysis," Bolt Beranek and Newman Inc., Cambridge, Mass., April 1974.
8. A.V. Oppenheim, R.W. Schafer and T.G. Stockham, "Nonlinear filtering of multiplied and convolved signals," *Proc. IEEE*, vol. 56, pp. 1264-1291, Aug. 1968.
9. J.L. Shanks, "Recursion filters for digital processing," *Geophysics*, vol. 32, pp. 33-51, Feb 1967.
10. R.E. Kalman, "Design of a self-optimizing control system," *Trans. ASME*, vol. 80, pp.468-478, Feb. 1958.

11. K. Steiglitz and L.E. McBride, "A technique for the identification of linear systems," IEEE Trans. Automatic Control, vol. AC-10, pp.461-464, Oct. 1965.
12. F. Brophy and A. C. Salazar, "Recursive digital filter synthesis in the time domain," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-22, pp.45-55, Feb.1974.
13. A. J. Berkhout, "Related properties of minimum-phase and zero-phase time functions," Geophysical Prospecting, vol.XXII, No.4, pp.683-709, 1974.
14. J.D. Markel and A.H. Gray, *Linear Prediction of Speech*, New York: Springer Verlag, 1976.
15. A.H. Gray and J.D. Markel, "Distance measures for speech processing," IEEE Trans. Acoust. Speech and Signal Processing, vol. ASSP-24, pp.380-391, Oct. 1976.
16. R.W. Schafer and L.R. Rabiner, "System for automatic formant analysis of voiced speech," J. Acoust. Soc. Amer., vol.47, pp.634-648, 1970.

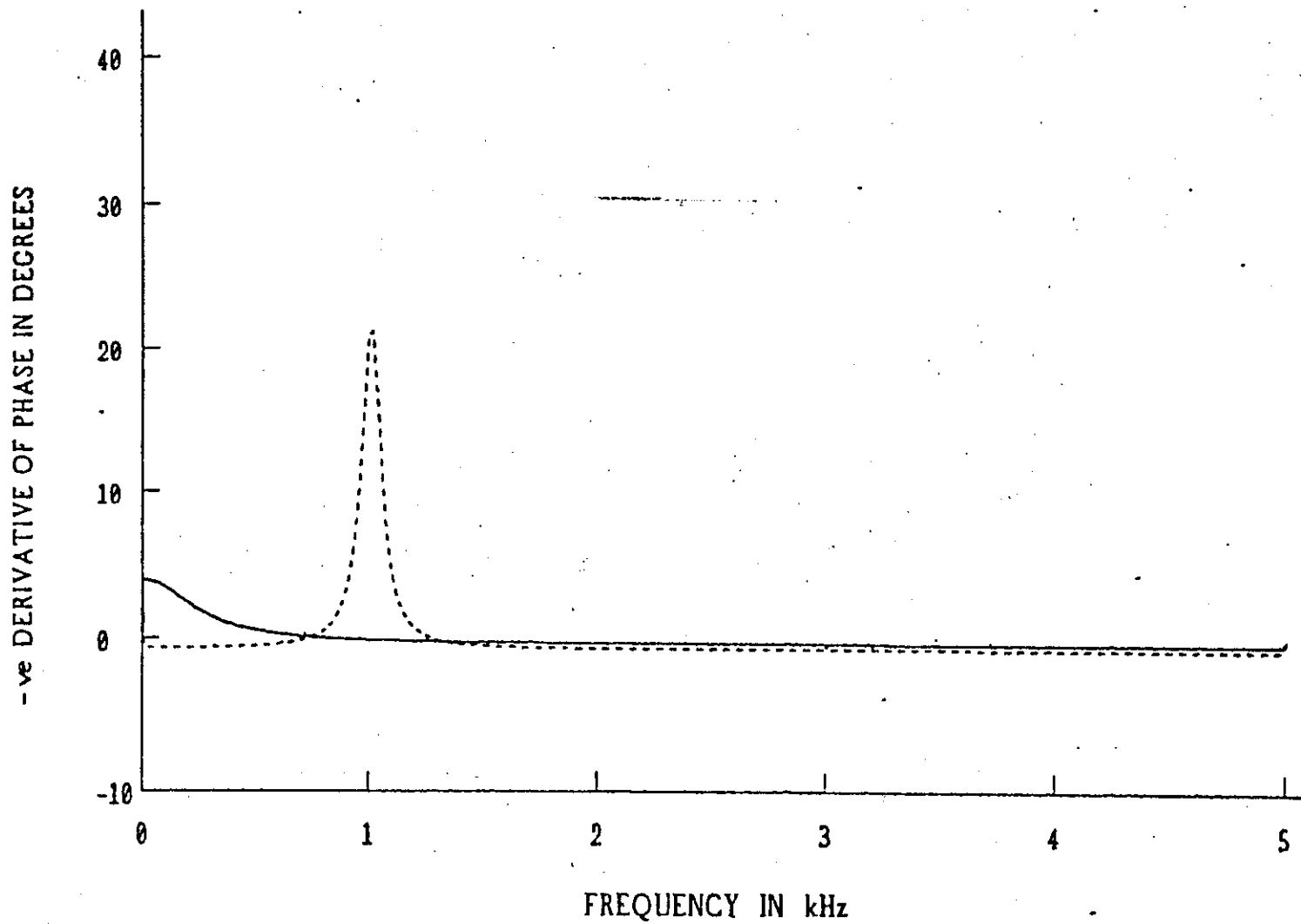


Fig.1. Negative derivative of phase spectra for typical all-pole filters.

(a) Solid curve: First order filter [$H(z)=1/(1-0.85z^{-1})$]

(b) Dotted curve: Second order filter [$H(z)=1/(1-1.57z^{-1}+0.94z^{-2})$]

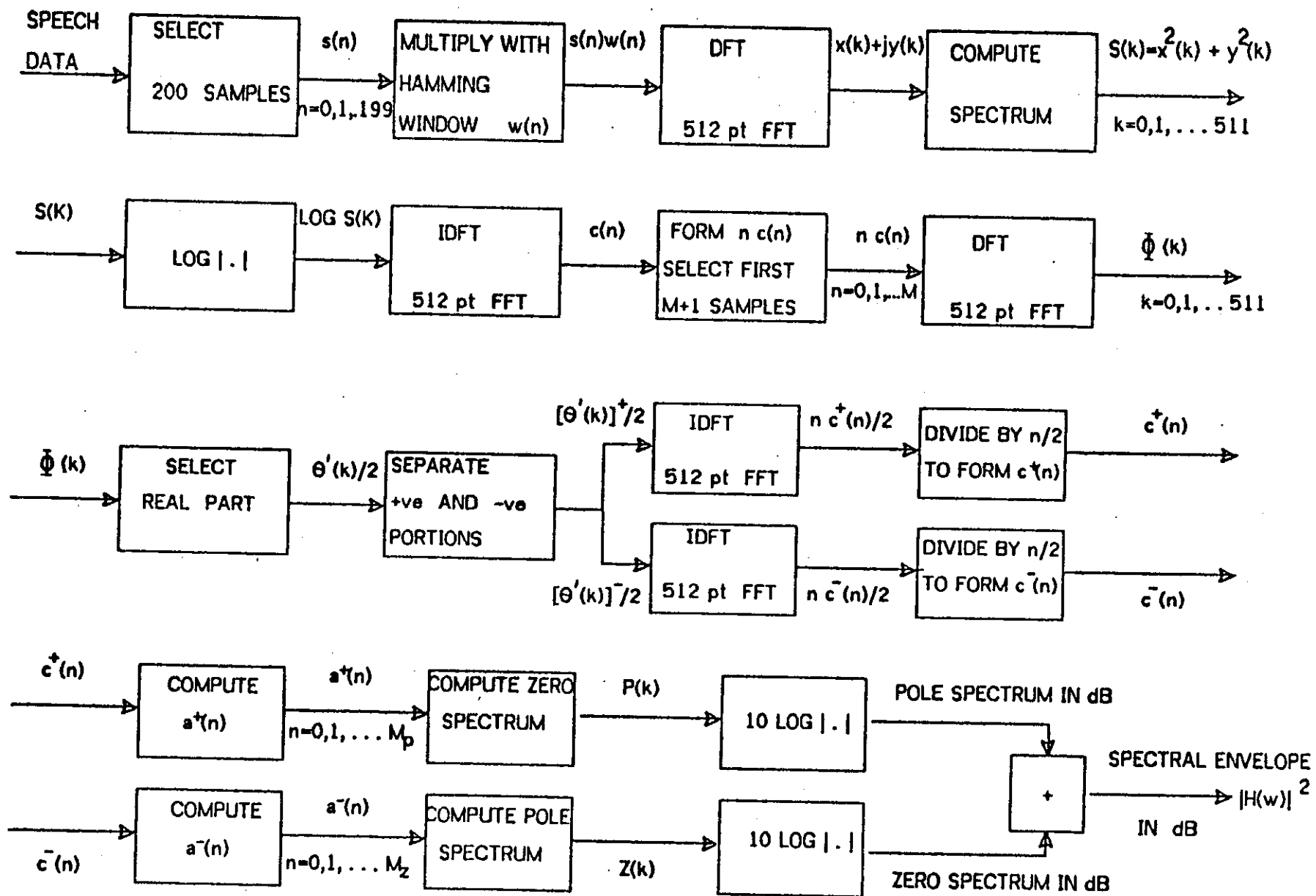


Fig.2 Block diagram showing computational steps for pole-zero decomposition

-ve DERIVATIVE OF PHASE SPECTRUM

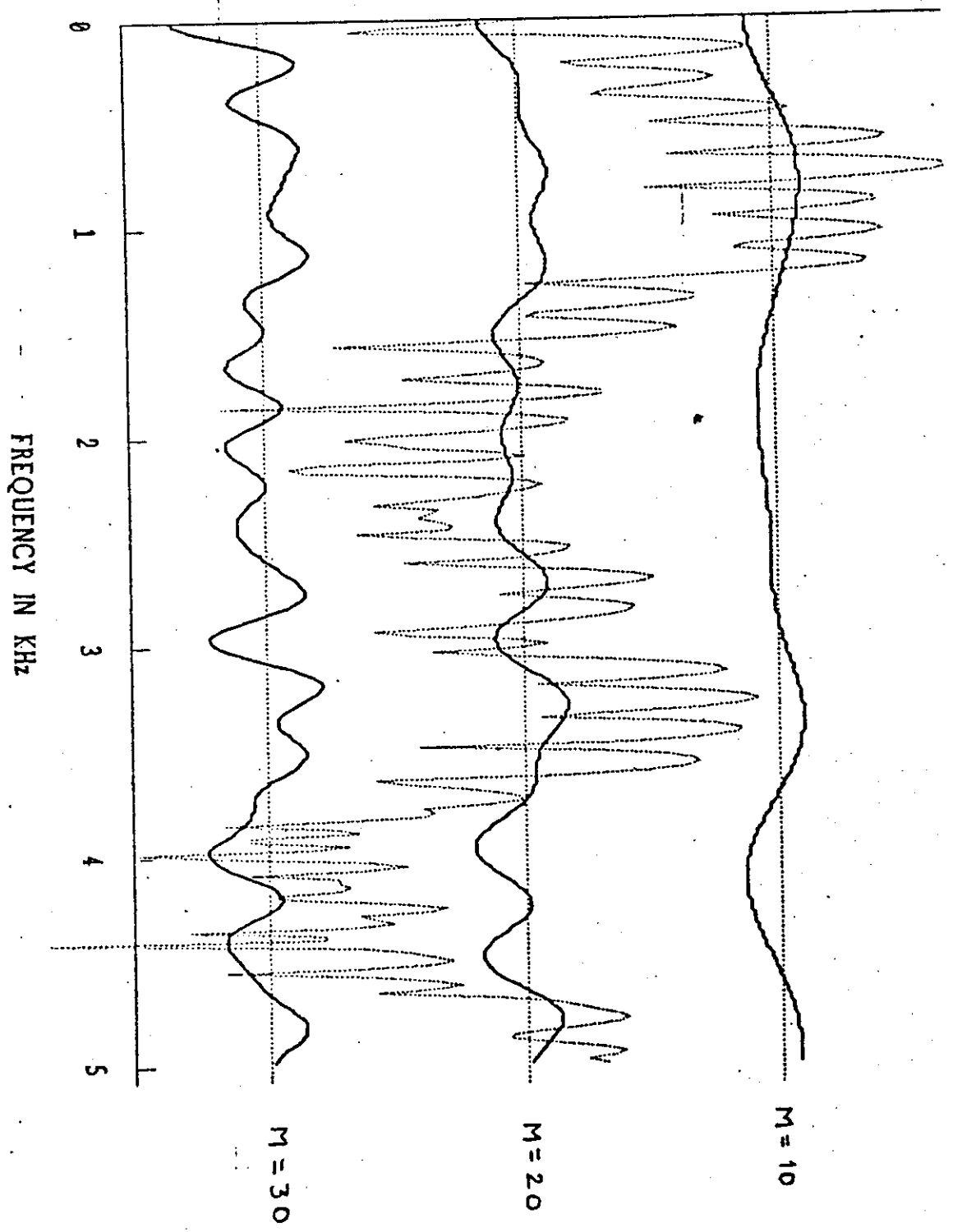


Fig.3. Negative derivative of phase spectra for a segment of voiced speech for different values of M .
(a) Solid curves: $M=10, M=20, M=30$. (b) Dotted curve: Short-time spectrum

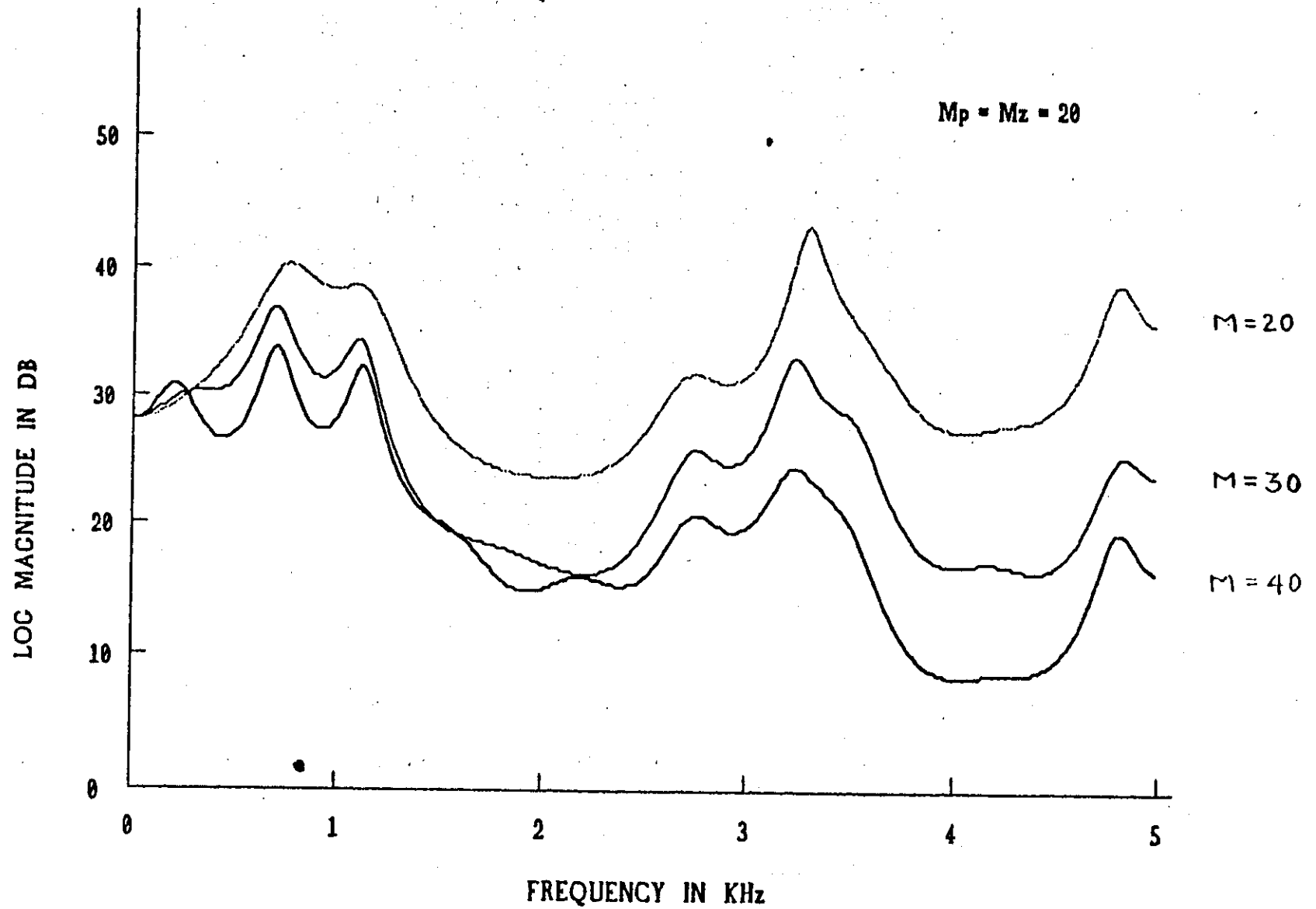


Fig.4. Pole spectra for for different values of M for a segment of voiced speech.
 $(M_p = M_z = 20)$

(a) Dotted curve: M=20 (b) Thin solid curve: M=30 (c) Thick solid curve: M=40

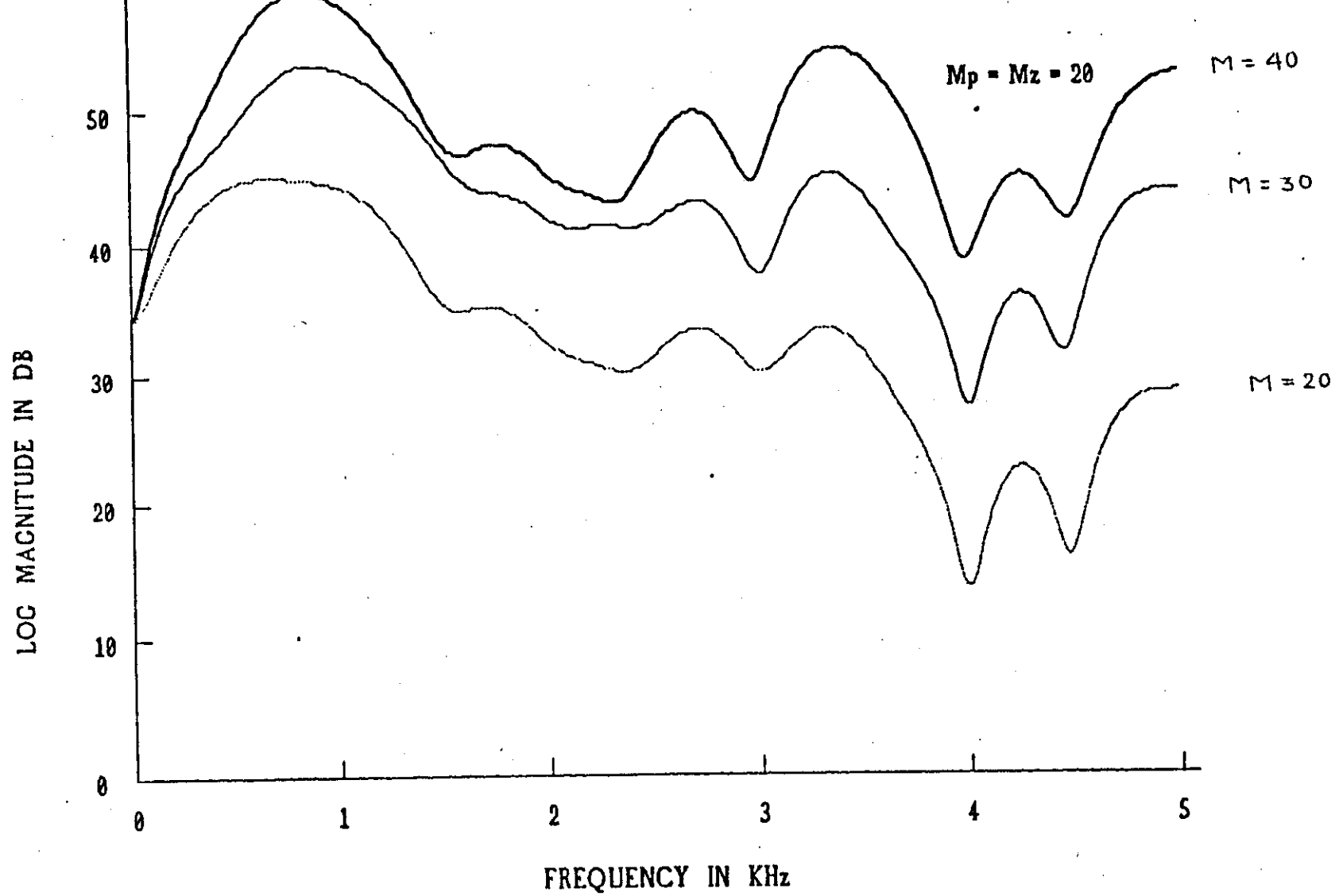


Fig.5. Zero spectra corresponding to Fig.4.

(a) Dotted curve: $M=20$ (b) Thin solid curve: $M=30$ (c) Thick solid curve: $M=40$

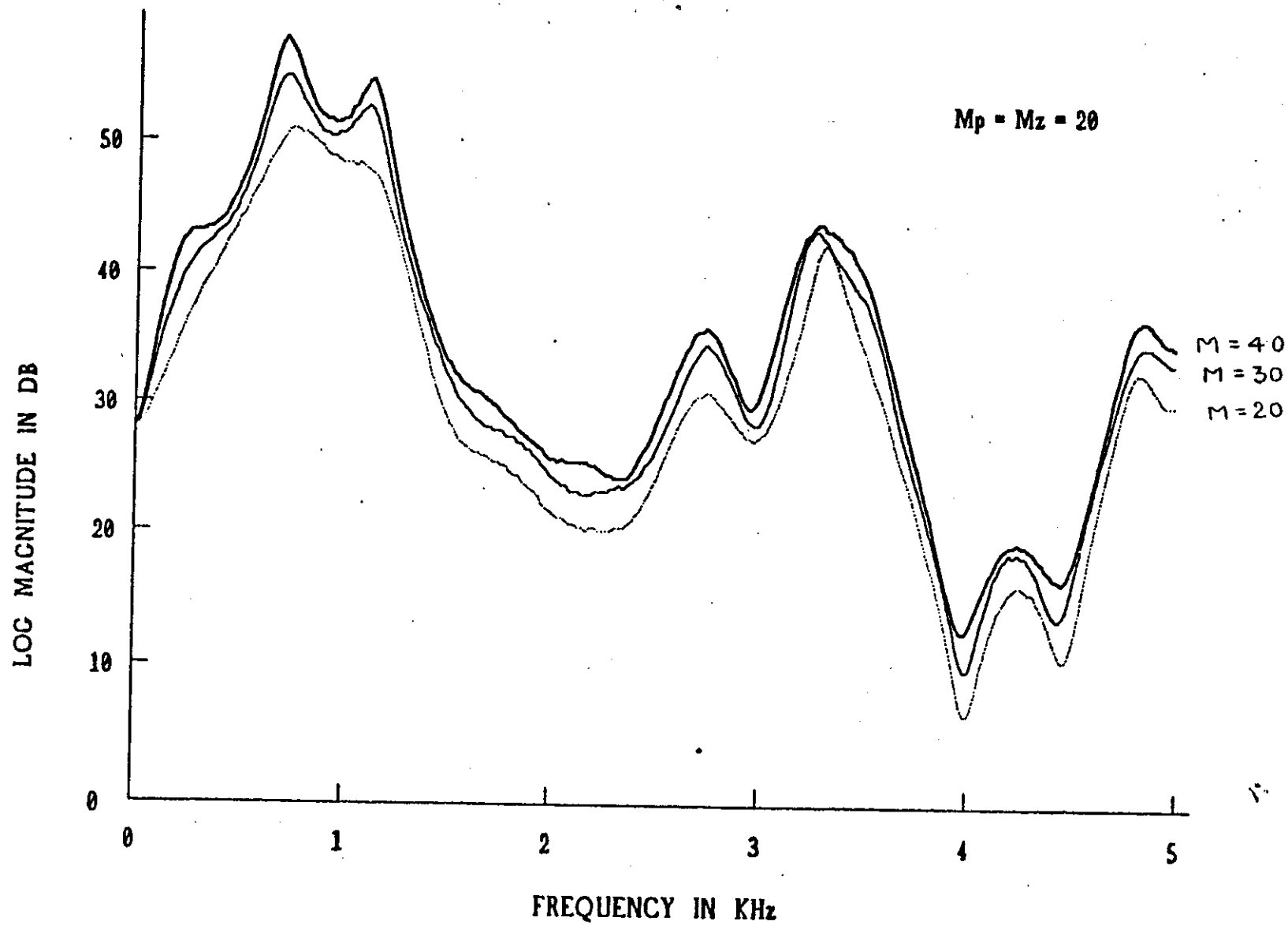


Fig.6. Pole-zero spectra for different values of M , obtained by adding the corresponding pole and zero spectra in Figs.4 and 5.

(a) Dotted curve: $M=20$ (b) Thin solid curve: $M=30$ (c) Thick solid curve: $M=40$

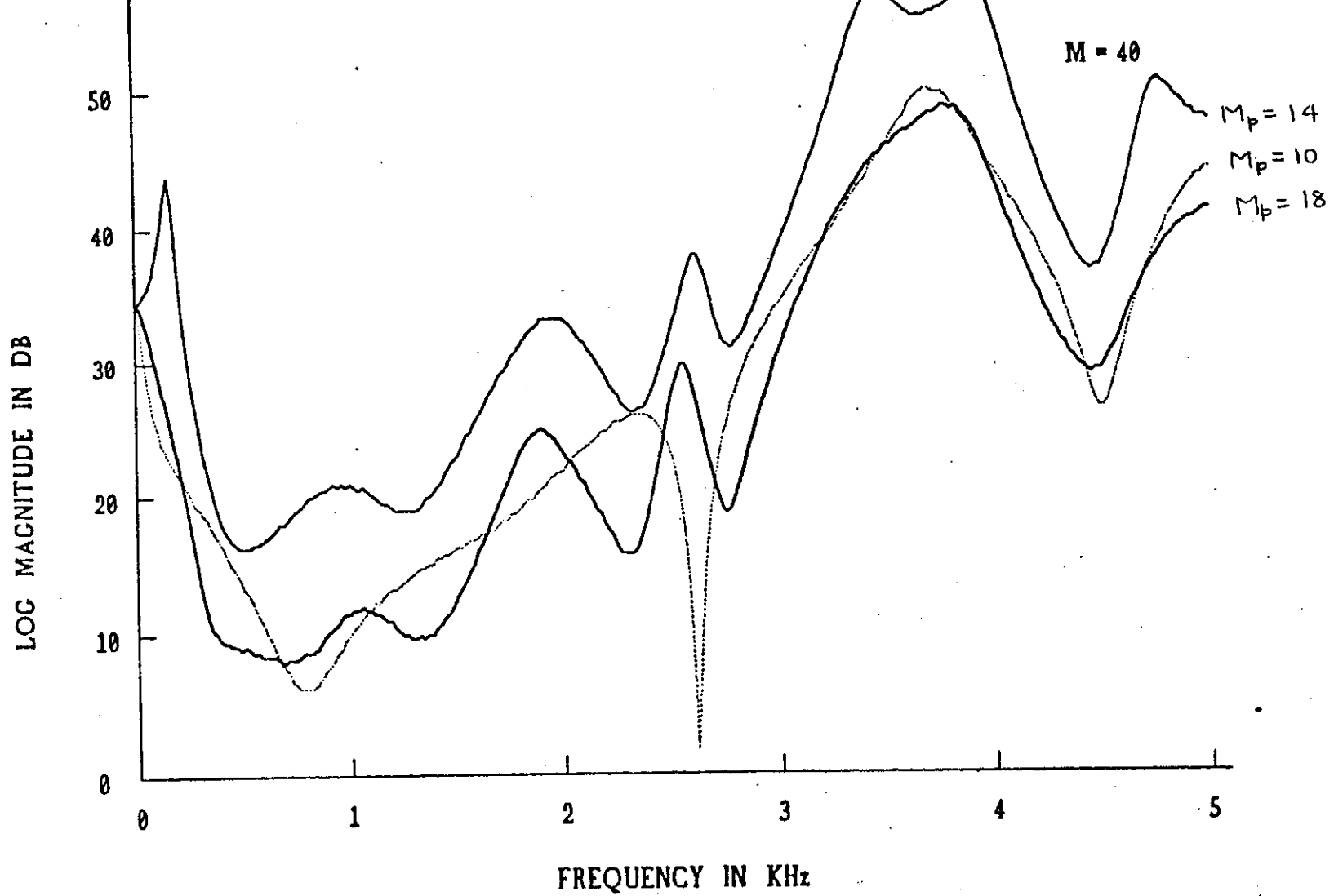


Fig.7. Pole-zero spectra for different values of $M_p (=M_z)$ for a segment of voiced fricative ($M=40$).

(a) Dotted curve: $M_p=10$. (b) Thin solid curve: $M_p=14$. (c) Thick solid curve: $M_p=18$.

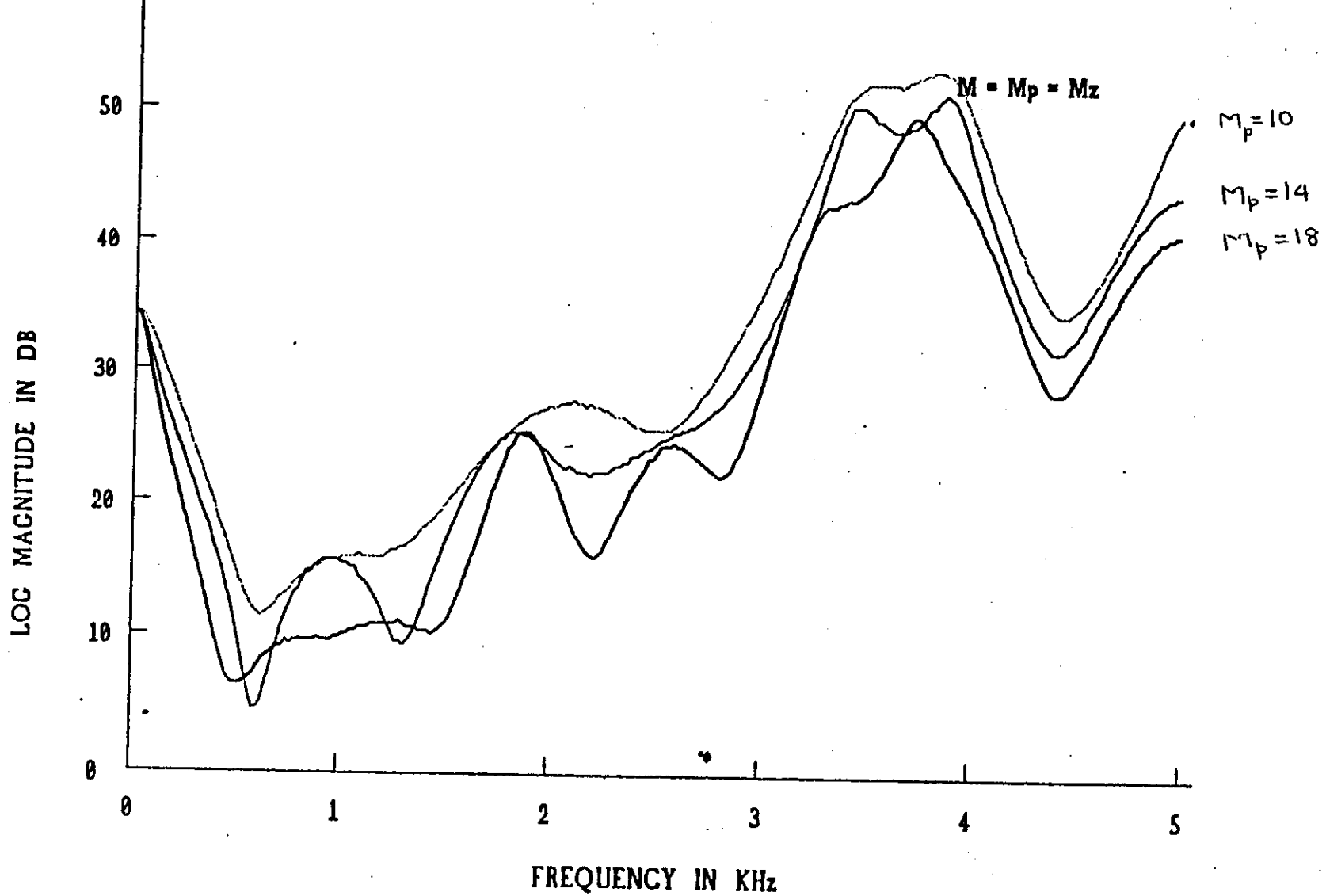


Fig.8. Pole-zero spectra for different values of $M_p (=M_z=M)$ for a segment of voiced fricative.

(a) Dotted curve: $M_p=10$ (b) Thin solid curve: $M_p=14$ (c) Thick solid curve: $M_p=18$

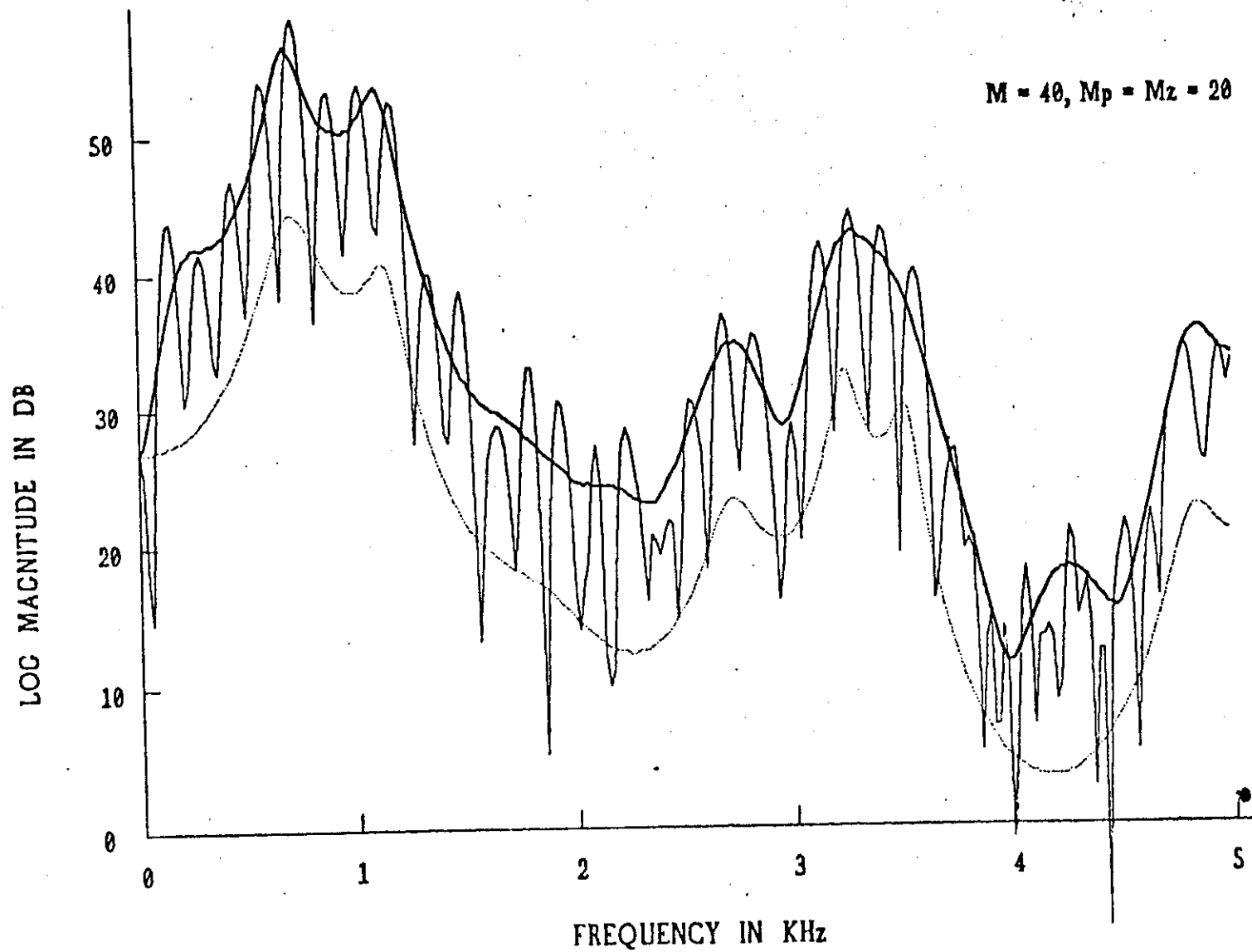


Fig.9. Pole-zero spectrum (thick solid curve) for a segment of diphthong in the word NINE. Short-time spectrum (thin solid curve) and 20-th order LP smoothed spectrum (dotted curve) are also shown for comparison.

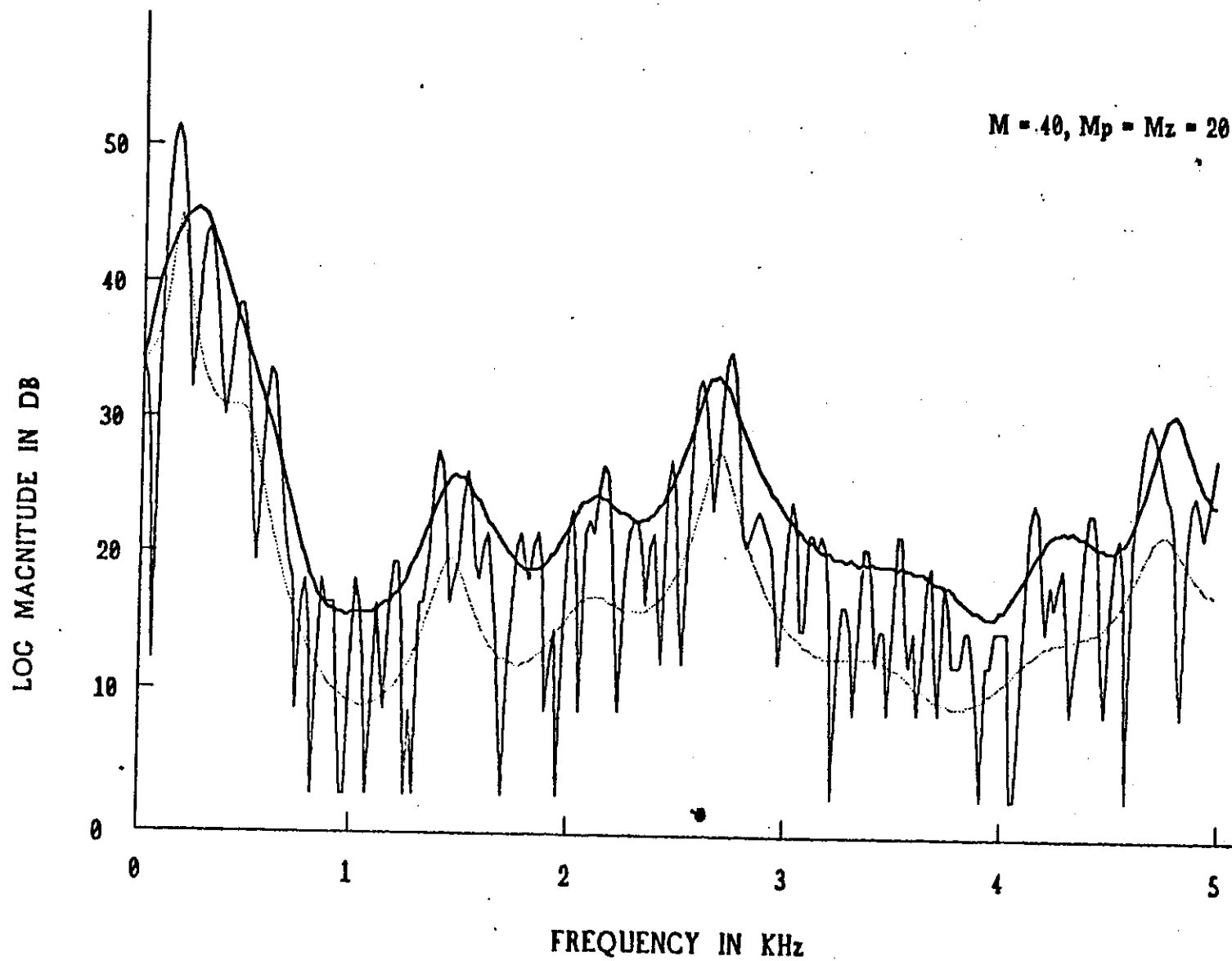


Fig.10. Pole-zero spectrum (thick solid curve) for a segment of nasal in the word NINE. Short-time spectrum (thin solid curve) and 20-th order LP smoothed spectrum (dotted curve) are also shown for comparison.

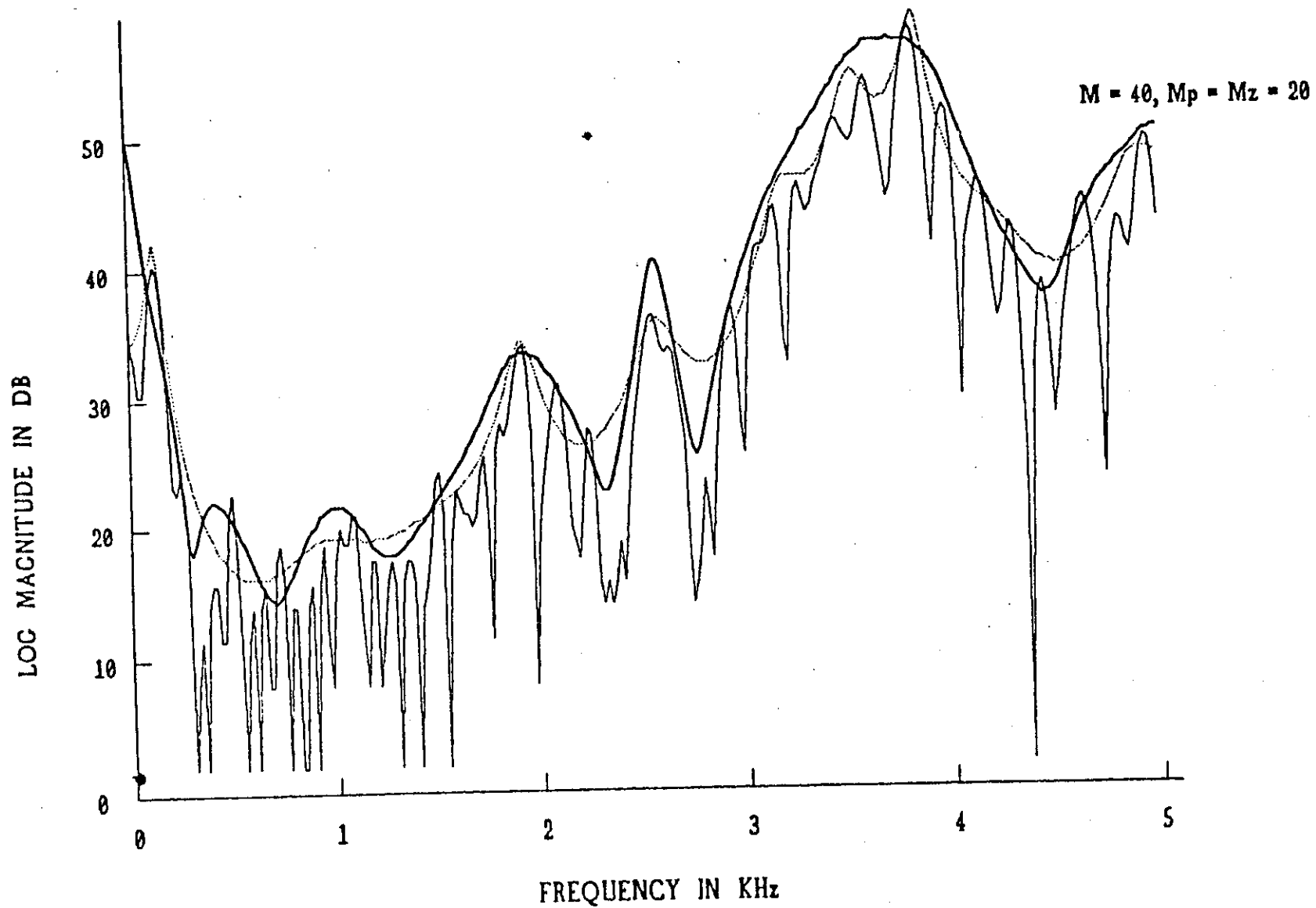


Fig.11. Pole-zero spectrum (thick solid curve) for a segment of voiced fricative in the word ZERO. Short-time spectrum (thin solid curve) and 20-th order LP smoothed spectrum (dotted curve) are also shown for comparison.

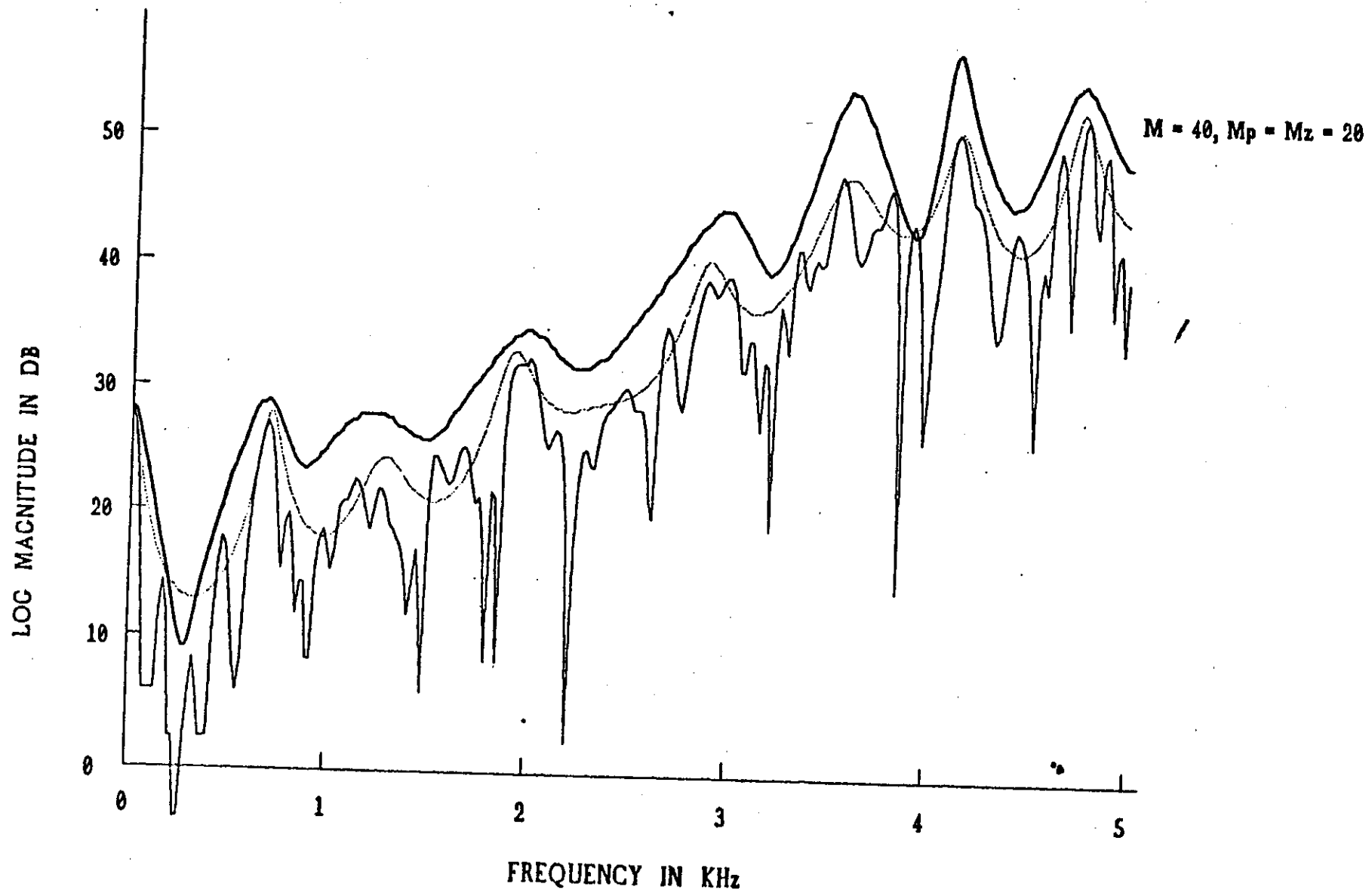


Fig.12. Pole-zero spectrum (thick solid curve) for a segment of unvoiced fricative in the word SIX. Short-time spectrum (thin solid curve) and 20-th order LP smoothed spectrum (dotted curve) are also shown for comparison.

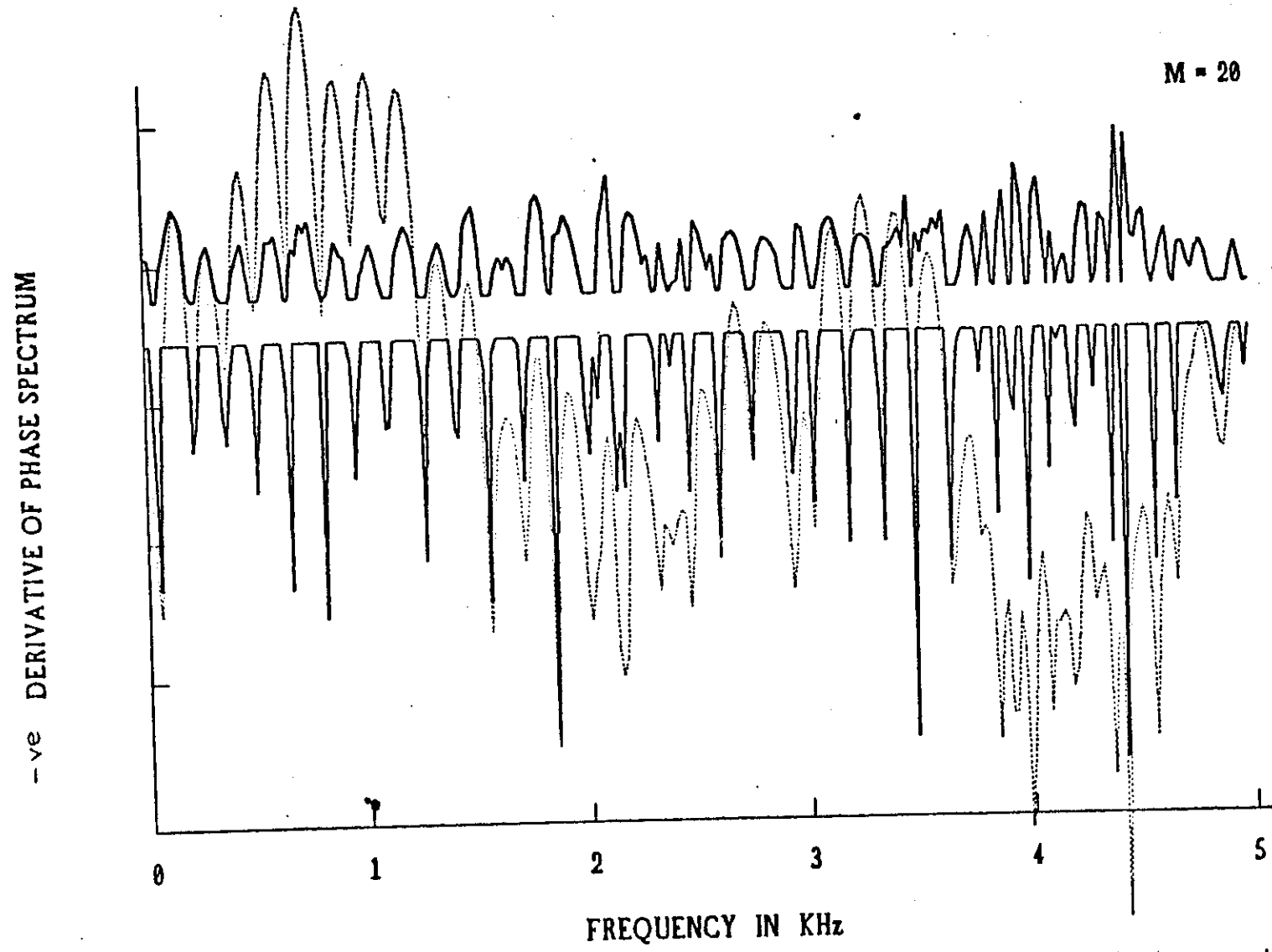


Fig.13. Negative derivative of phase spectrum for a voiced segment computed from equation (33).

(a) Upper thick solid curve: pole part (b) Lower thick solid curve: zero part (c) Dotted curve: Short-time spectrum of the voiced segment