Signal Processing 22 (1991) 259-267
Elsevier

Speech processing using group delay functions

Hema A. Murthy and B. Yegnanarayana
Department of Computer Science and Engineering, Indian Institute of Technology, Madras-600 036, India.

Received 8 December 1989
Revised 9 April 1990 and 8 August 1990

Abstract. In this paper we demonstrate the feasibility of processing the Fourier transform (FT) phase of a speeech signal to derive the smooth log magnitude spectrum corresponding to the vocal tract system. We exploit the additive property of the group delay function (negative derivative of the FT phase) to process the FT phase. We show that the rapid fluctuations in the log magnitude spectrum and the group delay function are caused by the zeroes of the z-transform of the excitation components of the speech signal. Zeroes close to the unit circle in the z-plane produce large amplitude spikes in the group delay function and mask the group delay information corresponding to the vocal tract system. We propose a technique to extract the vocal tract system component of the group delay function by using the spectral properties of the excitation signal.

Zusammenfassung. In diesem Beitrag zeigen wir daß aus den Phasen Fourier-Transformierten eines Sprachsignals ein geglättetes logarithmiertes Betragsspektrum gewonnen werden kann, das dem Sprachtrakt-Betragsfrequenzgang entspricht. Bei der Bearbeitung der Phasen des Spektrums wird die Additivitätseigenschaft der Gruppenlaufzeit (negative Ableitung des Phasengangs) ausgenutzt. Wir zeigen daß die schnellen Schwankungen im logarithmierten Betragsspektrum und der Gruppenlaufzeit durch die Nullstellen der Z-Transformierten der Sprachsignal-Erregung verursacht werden. Nullstellen in der Nähe des Einheitskreises der Z-Ebene rufen hohe Spitzen in der Gruppenlaufzeit-Funktion hervor und verdecken die Information, die in der Gruppenlaufzeit über den Sprachtrakt enthalten ist. Wir schlagen eine Methode vor, die es erlaubt, durch Ausnutzung der spektralen Eigenschaften des Erregungssignals aus der Gruppenlaufzeit die Sprachtrakt-Komponente zu extrahieren.

Résumé. Pour obtenir des informations pertinentes sur le signal de parole, on peut utiliser la phase, en exploitant les propriétés d'additivité et de haute résolution du temps de propagation de groupe (dérivée négative de la phase de la transformée de Fourier). On montre que les variations rapides du logarithme du spectre d'amplitude et du temps de propagation de groupe sont dues aux zéros de la transformée en Z de l'excitation. Les zéros, voisins du cercle unité, engendrent les pics d'amplitude élevés du temps de propagation de groupe et masquent les informations temporelles liées au conduit vocal. Pour estimer le temps de propagation de groupe, on propose une méthode ayant pour base les propriétés spectrales du signal d'excitation. Son comportement est évalué à l'aide d'une série d'expériences portant sur une gamme variée de signaux d'excitation. On montre que cette technique est bien adaptée au traitement des signaux de parole, surtout s'ils sont perturbés par des bruits additifs.

Keywords. Fourier transform phase, group delay functions, speech processing, formants.

1. Introduction

In the Fourier analysis of real data we have both magnitude and phase components, although the phase part is seldom used for parameter extraction. The phase spectrum of a signal appears to be noisy and difficult to process because it is available in a wrapped form (confined to the interval $\pm\pi$). But a significant feature of the phase function of a cascade of resonators is that the component phase spectra are additive, unlike that of the magnitude spectrum where the component magnitude spectra are multiplicative. In this paper we suggest a procedure to process the phase without destroying its additive property.

Our studies have shown that the FT phase is as important as the FT magnitude, and the relation between them can be explained through group delay functions [7]. The additive property of the phase is retained in the group delay function.

Another advantage of computing the group delay function rather than the phase spectrum is that the group delay function can be computed directly from the time domain signal without having to compute the unwrapped phase [4].

In the group delay domain the vocal tract and the excitation components are additive. The group delay function of speech is however difficult to process due to the presence of high amplitude positive and negative spikes corresponding to the spectral fine structure. These peaks are contributed by the zeroes close to the unit circle in the z-domain. In this paper we present a new method for extracting the group delay function corresponding to the vocal tract spectrum. In Section 2 we briefly discuss the properties of the group delay functions. In Section 3 we derive a modified group delay function corresponding to the vocal tract system. In Section 4 we study the effect of various parameters on the proposed method of deriving the smoothed magnitude response from the FT phase.

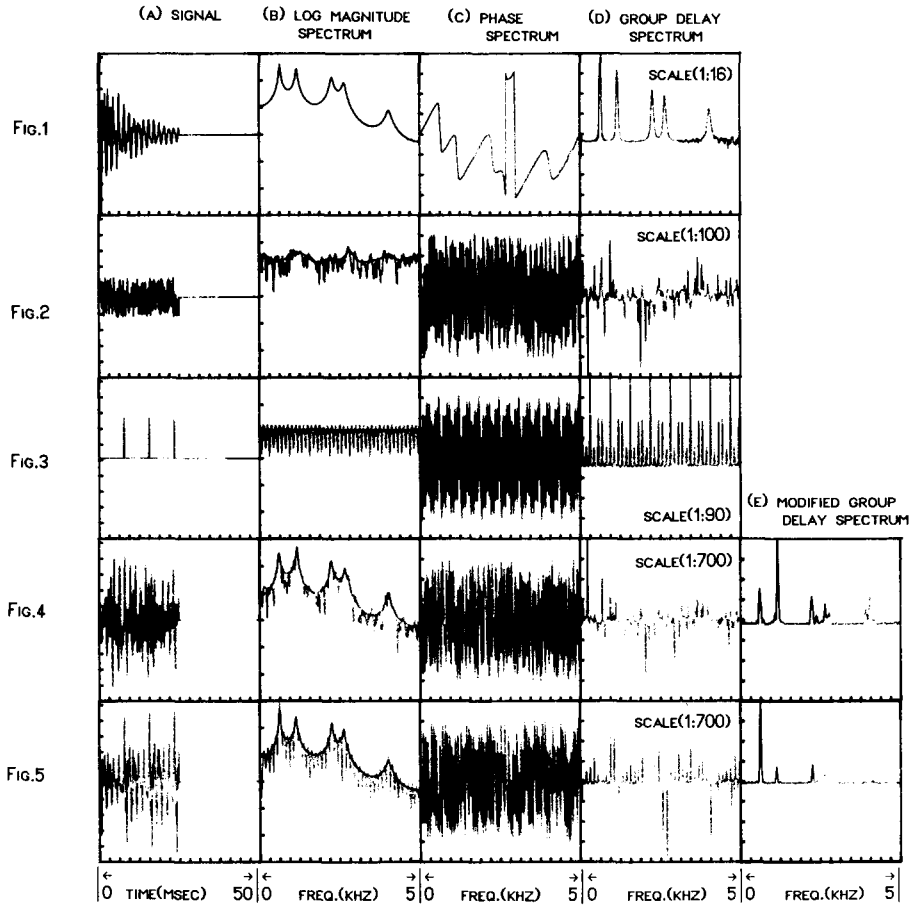

Fig. 1. Illustration of properties of the standard group delay function for the impulse response of an all-pole system. Fig. 2. Illustration of properties of the standard group delay function for a random noise sequence. Fig. 3. Illustration of properties of the standard group delay function for a train of impulses separated by a pitch period ($p = 80$). Fig. 4. Illustration of properties of the modified group delay function for a signal generated by exciting all-pole system with a random noise sequence, i.e., convolution of the signals in Figs. 1(A) and 2(A). Note that Fig. 4(E) is generated by multiplying the signal in Fig. 4(D) with an estimate of the excitation magnitude spectrum in Fig. 2(B). Fig. 5. Illustration of properties of the modified group delay function for a signal generated by exciting all-pole system with a periodic impulse train, i.e., convolution of the signals in Figs. 1(A) and 3(A). Note that Fig. 5(E) is generated by multiplying the signal in Fig. 5(D) with an estimate of the excitation magnitude spectrum in Fig. 3(B).

2. Theory and properties of group delay functions

In the theoretical discussion that follows initially we use continuous time and frequency variables and express the transfer function in terms of the Laplace transform. This helps us to visualize the resonance behaviour of the group delay function analytically. Later we use digital signals and the z-plane for the computation and discussion of the technique.

To explain the principle of the method, we consider a cascade of M resonators. The frequency response of the overall filter is given by

$$H(\omega)=\prod_{i=1}^{M}\frac{1}{(\alpha_i^2+\beta_i^2-\omega^2-2j\omega\alpha_i)}, \tag{1}$$

where $(\alpha_i \pm j\beta_i)$ is the complex pair of poles of the ith resonator. The magnitude spectrum is given by

$$|H(\omega)|^2=\prod_{i=1}^{M}\frac{1}{[(\alpha_i^2+\beta_i^2-\omega^2)^2+4\omega^2\alpha_i^2]}, \tag{2}$$

and the phase spectrum is given by

$$\theta(\omega)=\underline{/H}(\omega)=\sum_{i=1}^{M}\tan^{-1}\frac{2\omega\alpha_i}{\alpha_i^2+\beta_i^2-\omega^2}. \tag{3}$$

It is well-known that the magnitude of an individual resonator has a peak at $\omega^2=\beta_i^2-\alpha_i^2$ and a half-power bandwidth of α_i. We now consider the negative derivative of the phase spectrum (or group delay function)

$$\tau(\omega)=-\frac{d\theta(\omega)}{d\omega}=\sum_{i=1}^{M}\frac{2\alpha_i(\alpha_i^2+\beta_i^2+\omega^2)}{(\alpha_i^2+\beta_i^2-\omega^2)^2+4\omega^2\alpha_i^2}. \tag{4}$$

It was shown in [6] that around the resonance frequency $\omega_i^2=\beta_i^2-\alpha_i^2$ the group delay function behaves like a squared magnitude response. The response due to each resonator approaches zero asymptotically for ω away from the resonance frequency. The overall group delay function is a summation of the group delay functions due to individual resonators as can be seen from Fig. 1(D). Figure 1(A) shows the windowed impulse response of a 10th order all-pole filter. Figures 1(B, C, D) show the corresponding magnitude, phase and group delay spectra. Note that the group delay function (Fig. 1(D)) has sharp peaks around the resonances due to the squared magnitude behaviour and has very small values in between two resonance peaks due to the asymptotic behaviour for frequencies away from the resonance frequency.

It was shown in [1] that the digitally computed group delay functions accurately represent the signal information as long as the roots of the signal z-transform are not too close to the unit circle in the z-plane. It was noticed that adequate sampling based on the Nyquist criterion in the time domain does not necessarily result in proper sampling in the group delay domain.

3. Basis for the proposed method: Modified group delay function

In digital processing of speech signals, the vocal tract system and the excitation contribute to the envelope and the fine structure, respectively, of the spectrum. Techniques used to extract resonances from the FT magnitude try to capture the spectral envelope and disregard the fine structure. Similarly, to derive the vocal tract characteristics from the group delay function, the component due to spectral fine structure must be deemphasized. Zeroes close to the unit circle manifest as spikes in the group delay function. The strengths of these spikes depends upon the proximity of the zeroes to the unit circle. The polarity of the spikes depends on whether the zero lies inside or outside the unit circle. These spikes form a significant part of the fine structure and their effect cannot be eliminated by normal smoothing techniques. In our previous attempts [2], the speech signal was modified prior to the group delay computation to reduce the effect of the spikes in the group delay domain. Here we suggest a method for reducing the contribution of the fine structure to the group delay function by modifying the group delay function derived directly from the time domain signal. This

modification is based on the conjecture that the spikes in the group delay function are caused by zeroes close to the unit circle. Our initial attempts to compensate for the zeroes involved modifying the expression for computing the group delay function in an ad hoc manner which is reported in [3]. We now substantiate this conjecture with both a theoretical analysis and experimental results, and suggest a modification which does not involve any empirical parameters.

Since speech signal can be characterized as the response of an all-pole filter to an excitation either from a periodic train of impulses or from a random noise sequence, the z-transform of the system generating the speech signal can be written as

$$H(z)=\frac{N(z)}{D(z)}. \tag{5}$$

The numerator polynomial $N(z)$ corresponds to the contribution by the excitation, and the denominator polynomial $D(z)$ corresponds to the contribution by the poles of the vocal tract system. The frequency response of $H(z)$ is given by

$$H(\omega)=\frac{N(\omega)}{D(\omega)}, \tag{6}$$

where $H(\omega)$, $N(\omega)$ and $D(\omega)$ are obtained by evaluating the corresponding polynomials on the unit circle in the z-plane.

The group delay (negative derivative of the phase) function of $H(\omega)$ is given by

$$\tau(\omega)=\tau_N(\omega)-\tau_D(\omega), \tag{7}$$

where $\tau_N(\omega)$ and $\tau_D(\omega)$ are the group delay functions corresponding to $N(\omega)$ and $D(\omega)$. We have already discussed the shape and properties of $-\tau_D(\omega)$ through (4). Although it is difficult to derive an analytical expression for $\tau_N(\omega)$, we can study its behaviour in terms of the characteristics of excitation signals. Since $N(z)$ corresponds to the z-transform of the excitation signal, the zeroes of $N(z)$ close to the unit circle produce large amplitude spikes in $\tau_N(\omega)$. The polarity of the spikes depends on whether the zeroes are lying inside or outside the unit circle in the z-plane. Figures 2 and 3 illustrate the behaviour of $\tau_N(\omega)$ for a random noise sequence and impulse train, respectively. Note that the log magnitude spectra (Figs. 2(B) and 3(B)) have nearly a flat spectral envelope with rapid fluctuations superimposed on it due to zeroes close to the unit circle. The group delay functions (Figs. 2(D) and 3(D)) have large fluctuations around zero. The large positive and negative spikes of $\tau_N(\omega)$ mask the details of the resonance peaks due to $-\tau_D(\omega)$ in the combined response $\tau(\omega)$. This is illustrated in Figs. 4 and 5. The signal in Fig. 4 corresponds to a windowed version of the signal generated by convolving the random noise sequence (Fig. 2(A)) with the impulse response (Fig. 1(A)) of an all-pole system. The group delay function (Fig. 4(D)), which is simply the sum of the plots of Fig. 1(D) and 2(D) shows that the resonance peaks are indeed masked by the large amplitude spikes. Note that the vertical scales in Figs. 1(D) and 2(D) are different, the peak amplitudes in Fig. 2(D) being very much larger than the amplitudes in Fig. 1(D). Similar behaviour is observed in Fig. 5, where the signal is a windowed version of the signal obtained by convolving the impulse train in Fig. 3(A) with the impulse response (Fig. 1(A)) of an all-pole system.

The equation for $\tau(\omega)$ can be written as [4]

$$\tau(\omega)=\frac{X_R(\omega)Y_R(\omega)+X_I(\omega)Y_I(\omega)}{|X(\omega)|^2}, \tag{8}$$

where $X(\omega)$ and $Y(\omega)$ are the Fourier transforms of the discrete-time signals $x(n)$ and $y(n)=nx(n)$, and the subscripts R and I stand for the real and imaginary parts, respectively. In the expression for computing $\tau_N(\omega)$, $|N(\omega)|^2$ appears in the denominator. Small values of $|N(\omega)|^2$ at frequencies near zeroes of $N(\omega)$ contribute to the large amplitude spikes. For computing $\tau_D(\omega)$, the term $|D(\omega)|^2$ appears in the denominator. Since $D(z)$ has all the roots well within the unit circle, $|D(\omega)|^2$ will not have very small values as in $|N(\omega)|^2$. Hence $\tau_D(\omega)$ will not have large amplitude spikes as in $\tau_N(\omega)$. The combined group delay

function is now given by

$$\tau(\omega) = \tau_N(\omega) - \tau_D(\omega) = \frac{\alpha_N(\omega)}{|N(\omega)|^2} - \frac{\alpha_D(\omega)}{|D(\omega)|^2}, \quad (9)$$

where $\alpha_N(\omega)$ and $\alpha_D(\omega)$ are the numerator terms of (8) for $\tau_N(\omega)$ and $\tau_D(\omega)$, respectively.

Suppose we multiply $\tau(\omega)$ with $|N(\omega)|^2$. Then the contribution due to the zeroes is significantly reduced. Since the envelope of $|N(\omega)|^2$ is nearly flat, the significant features (resonance peaks) of the second term will still shown up, with superimposed fluctuations of $|N(\omega)|^2$. The modified group delay function is given by

$$\tau_0(\omega) = \tau(\omega)|N(\omega)|^2 = \alpha_N(\omega) - \frac{\alpha_D(\omega)}{|D(\omega)|^2}|N(\omega)|^2. \quad (10)$$

In (10) the contribution of the first term $\alpha_N(\omega)$ should be small compared to the second term in order to emphasize the group delay component of the second term. That $\alpha_N(\omega)$ is small for a random noise sequence can be seen from Fig. 4(E), where the modified group delay function $\tau_0(\omega)$ is plotted for the signal in Fig. 4(A). Note that between two resonance peaks the value of $\tau_D(\omega)$ is nearly zero (as discussed earlier) due to the additive property of the group delay function. That is why the modified group delay function resembles the group delay function for the impulse response of the all-pole system as can be seen from Figs. 4(E) and 1(D). Note that the modified group delay function in Fig. 4(E) is obtained by multiplying the function in Fig. 4(D) with an estimate of the excitation spectrum in Fig. 2(B). Figure 5 illustrates similar results for the periodic impulse excitation. Later in the experiments we show that for a variety of excitation functions $\alpha_N(\omega)$ is small.

Therefore, the problem of determining the component due to the resonances is reduced to the estimation of the function $|N(\omega)|^2$. In practice $|N(\omega)|^2$ has to be estimated from the given signal. It is important to preserve the values of $|N(\omega)|^2$ around the zeroes so that it cancels the small values in the denominator of the first term in (9). Therefore $|N(\omega)|^2$ should retain all the sharp fluctuations of the log magnitude spectrum and should have a flat spectral envelope. We will show that the second condition is not as critical as the first one. An approximation $\hat{Z}(\omega)$ to $|N(\omega)|^2$ can be obtained by dividing the signal spectrum ($S(\omega) = |H(\omega)|^2$) with a cepstrally smoothed spectrum $V_c(\omega)$ [4, p. 519]. That is,

$$\hat{Z}(\omega) = \frac{S(\omega)}{V_c(\omega)}, \quad (11)$$

where $S(\omega)$ is the signal spectrum and $V_c(\omega)$ is the cepstrally smoothed spectrum of $S(\omega)$. Figures 4(E) and 5(E) show the results of processing the group delay function using an estimate $\hat{Z}(\omega)$ for $|N(\omega)|^2$ derived from a cepstrally smoothed spectrum of the signal. The figures show that we have indeed obtained a group delay function that is close to Fig. 1(D).

Since the modified group delay function roughly corresponds to the group delay function of an all-pole system, it is possible to derive the corresponding log magnitude spectrum using the minimum phase property. The algorithm involves computation of the linearly weighted cepstral coefficients from $\tau_0(\omega)$, followed by the computation of the cepstral coefficients and finally the log magnitude [7].

4. Effect of various parameters

While the group delay function has many interesting properties, its computation in the digital domain causes some problems. We have conducted a series of experiments to study the robustness of the proposed technique. The choice of the experiments is based upon the discussion given in an earlier paper [1] and our own experience with the use of group delay functions over the past several years. Composite signals of the form shown in (12) are used in these experiments. Each signal is obtained as the response of a five formant vocoder to a train of impulses separated by a pitch period. The amplitude of the impulses are 1, γ, γ^2, γ^3, ...

The composite signal is given by

$$y(n) = x(n) + \gamma x(n-p) + \gamma^2 x(n-2p) + \gamma^3 x(n-3p) + \cdots, \quad (12)$$

where $x(n)$ is the basic signal corresponding to the impulse response of the system. Taking the z-transform of the above equation we get

$$Y(z) = \frac{X(z)}{1-\gamma z^{-p}}. \quad (13)$$

This signal contains 5 pairs of complex conjugate pole pairs located inside the unit circle in the z-plane due to the basic signal. The distribution of zeroes and the number of zeroes are determined by the values γ and p, respectively. If $\gamma = 0$, we only have the basic signal. The choice of such a signal is justified because speech is indeed a type of composite signal. In the following experiments a particular parameter is varied, the modified group delay function and the corresponding log magnitude spectrum of the vocal tract system are computed. The performance is judged by comparing the modified group delay function with the true group delay function for the vocal tract system for synthetic signals. The smoothed log magnitude spectrum is also used to demonstrate the usefulness of the proposed method.

A few comments are given here to explain the organisation of the plots in our studies. For each case we have given the time domain signal usually of 256 samples, followed by the log magnitude spectrum of the signal. A 16th order LP spectrum is superimposed on the log magnitude spectrum. For synthetic signals the LP spectrum corresponds to the ideal log magnitude spectrum of the vocal tract system. Our main aim is to show that it is possible to process the Fourier transform phase through the group delay functions. Therefore in each figure the phase spectral plots are given to illustrate the complexity of the phase data due to wrapping. The complexity is reduced in the group delay function plot because the effects of wrapping are absent. However the effect of zeroes close to the unit circle mask the information about the vocal tract system. The complexity is further reduced in the modified group delay plot to bring out the features (like formants) of the vocal tract system. Finally a smoothed log magnitude spectrum is given corresponding to the modified group delay function. In all the figures vertical scale is not explicitly mentioned, since we are only looking at the features in the plots.

Experiment 1: Effects of various analysis parameters

We have considered the effect of each of the following parameters on the modified group delay function and the resulting smoothed log magnitude spectrum:

(a) Size of cepstral window to derive $\hat{Z}(\omega)$ in (11).
(b) Size and shape of the analysis window for the signal.
(c) Proximity of zeroes to the unit circle by varying γ in (12).
(d) Number of zeroes by varying p in (12).
(e) Proximity of resonances.

We have found that the modified group delay function is almost the same over a range 4 to 20 samples of the cepstral window used to derive $\hat{Z}(\omega)$ in (11). This shows that the size of the cepstral window is not very critical in the proposed method of deriving the smoothed magnitude spectrum. Our studies on the effects of the other parameters also show that the method is not critically dependent on the size and shape of the data window, the distribution of the zeroes due to excitation and the distribution of resonances of the system. It should be noted, however, that the limit on the resolution of the formants peaks is governed by the size of the data window, since our starting point is still the discrete Fourier transform of the given data for computation of the modified group delay function.

Experiment 2: Different types of excitation functions

So far we have only considered synthetic signals which correspond to the response of an all-pole system to a sequence of periodic impulses. In this experiment we compare the modified group delay functions derived from signals generated using

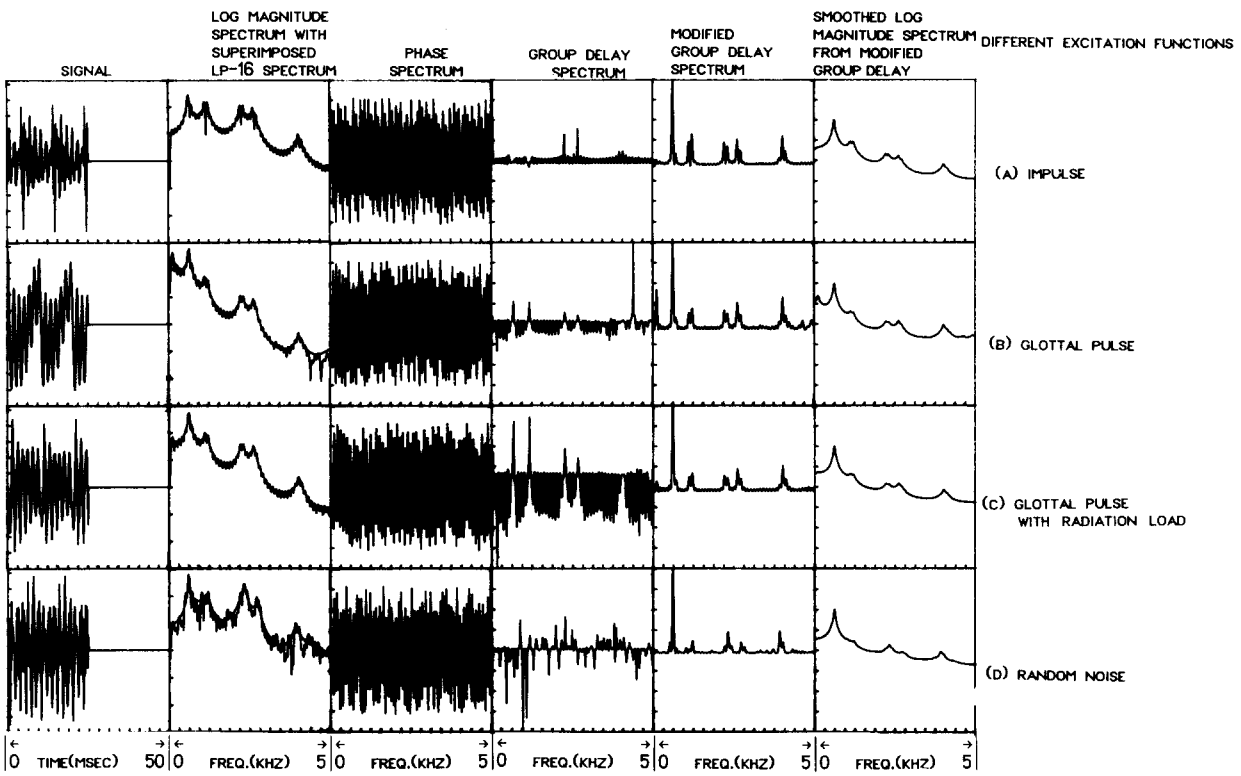

Fig. 6. Illustration of the effect of different excitation functions on the modified group delay function: (A) impulse, (B) glottal pulse, (C) glottal pulse with radiation load and (D) random noise.

four different excitation functions: (a) An impulse sequence separated by pitch period (100 samples); (b) Synthetic glottal pulse sequence as defined by Rosenberg [5, p. 103]; (c) Glottal pulse sequence with radiation load [5, p. 102] and (d) Uniformly distributed random noise. Figure 6 shows the results for the different excitation functions. We can see that the effect of these excitation functions on the modified group delay function is minimal. This is due to the fact that all the excitation functions are finite duration signals which introduce zeroes in the z-plane. The effect of these zeroes is reduced in the modified group delay function.

Experiment 3. Natural speech

In this experiment we consider different segments of natural speech. Figure 7 shows the plots for four consecutive segments of speech chosen arbitrarily from an all-voiced utterance. The results show that the formant information is preserved in the modified group delay function. Note that the resulting log magnitude spectra are derived without implying any model for the vocal tract system.

Experiment 4. Noisy speech data

In this experiment we consider an arbitrarily chosen segment of synthetic speech which is corrupted by additive white Gaussian noise. The signal-to-noise ratio (SNR) is progressively decreased. The effect on the modified group delay function is shown in Fig. 8. Notice that significant features are preserved even when the SNR is 0 dB. This point is also illustrated in Fig. 9 for natural speech. An important observation is that the spectral dynamic range has been restored to almost the original value as can be seen from the derived smoothed log magnitude function. These spectra also preserve the formant information as can be seen from the modified group delay function. Formant peaks at low SNR regions in the spectrum are obviously lost in most cases.

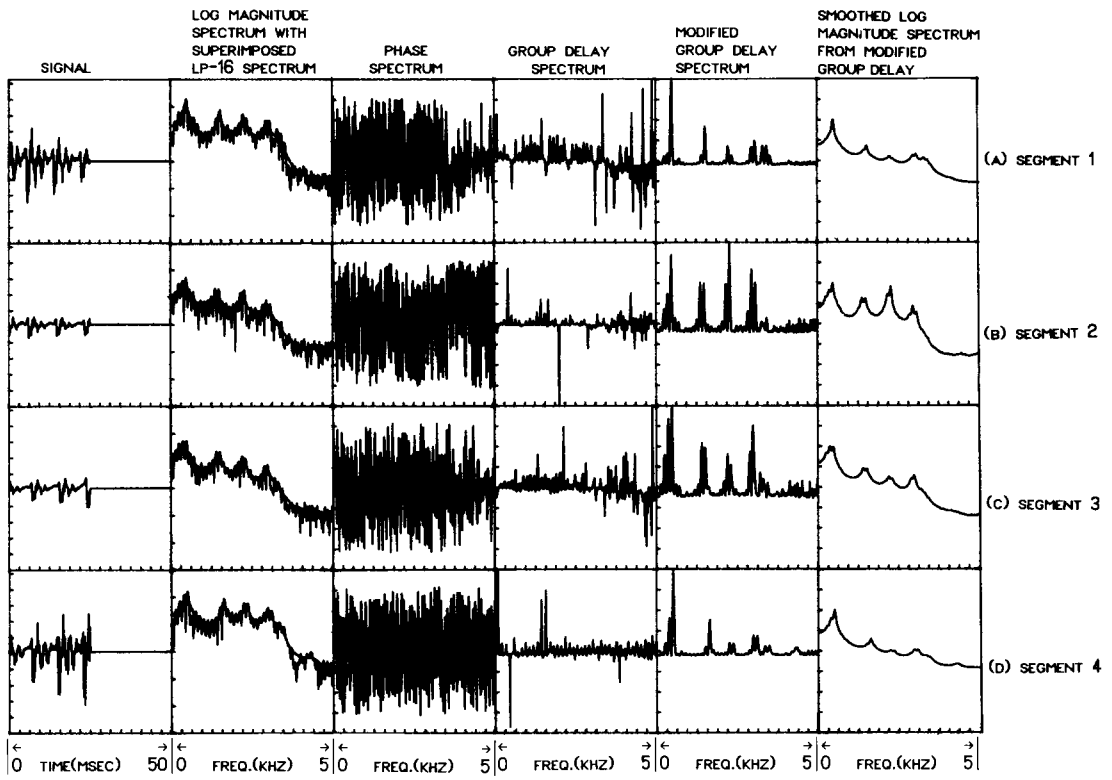

Fig. 7. Illustration of modified group delay functions for some segments of natural speech.

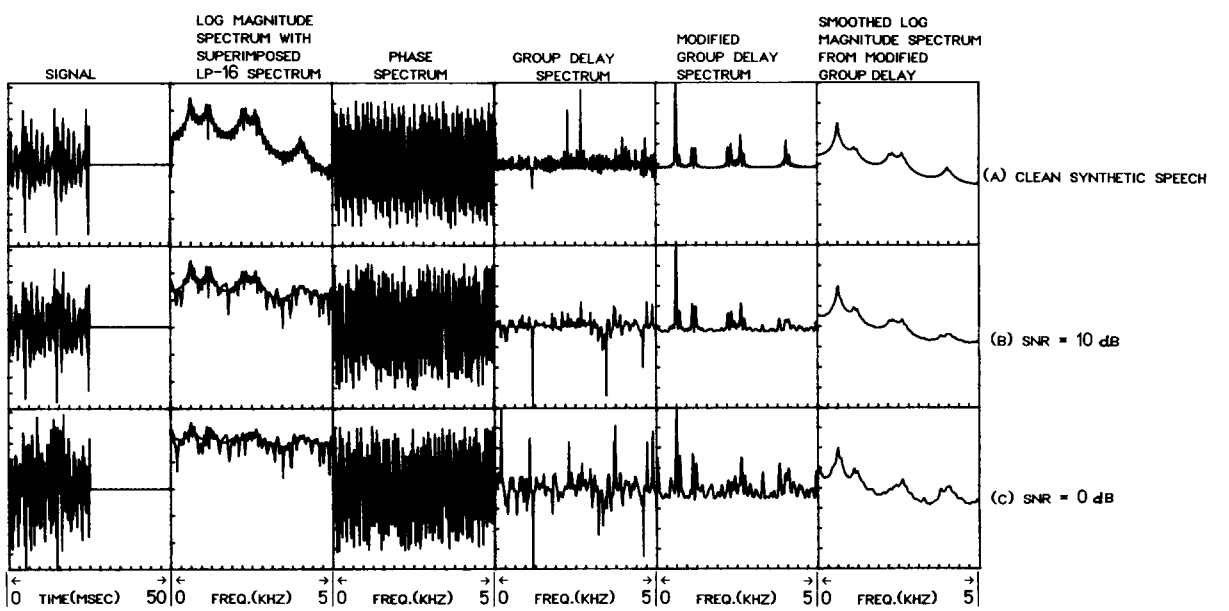

Fig. 8. Effect of noise on the modified delay function (synthetic speech): (A) clean signal, (B) SNR = 10 dB and (C) SNR = 0 dB

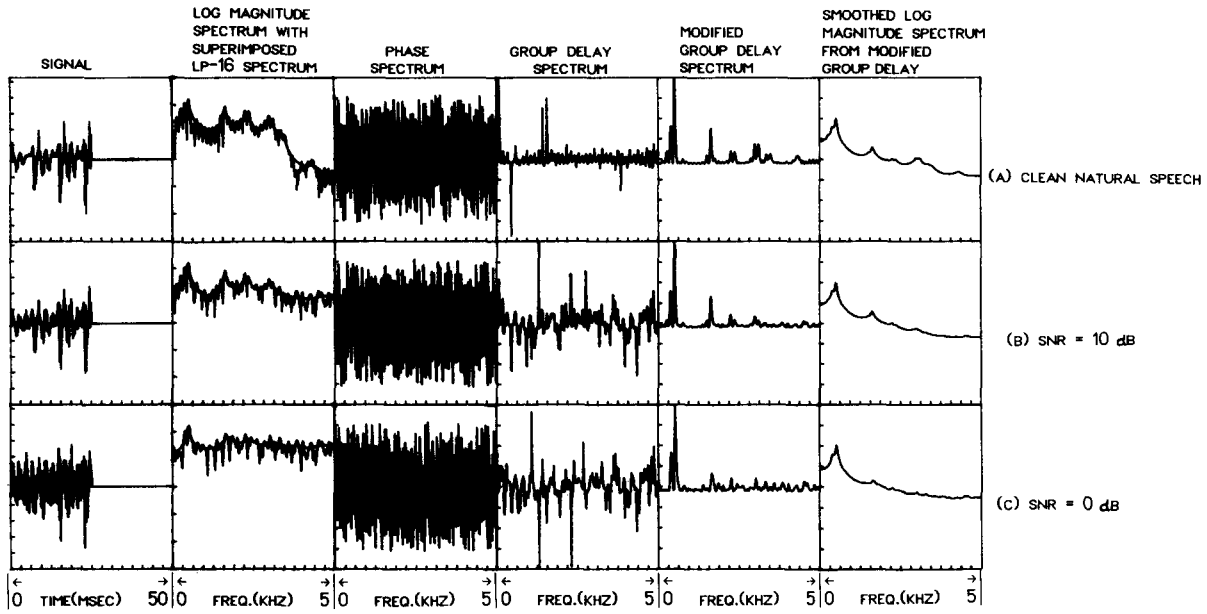

Fig. 9. Effect of noise on the modified group delay function (natural speech): (A) clean signal, (B) SNR = 10 dB and (C) SNR = 0 dB.

5. Conclusions

In this paper we have proposed a new technique for processing the Fourier transform phase spectrum of a speech signal. The standard phase spectrum is considered difficult to interpret due to the artifacts introduced by the zeroes of the z-transform of the excitation function. We have proposed a technique to process the phase in which the effect of these zeroes is significantly reduced. The main results of this study are

(1) The fluctuations caused by zeroes are reduced.
(2) The effects of time window functions are significantly reduced.
(3) The most striking result is that the spectral dynamic range appears to be restored for noisy speech in most cases.

References

[1] K.V. Madhu Murthy and B. Yegnanarayana, "Effectiveness of representation of signals through group delay functions", *Signal Processing*, Vol. 17, No. 2, June 1989, pp. 141-150.

[2] H.A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana, "Formant extraction from Fourier transform phase", *Proc. Internat. Conf. Acoust. Speech Signal Process.*, May 1989, pp. 484-487.

[3] H.A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana, "Formant extraction from phase using weighted group delay functions", *Electron. Lett.*, Vol. 25, No. 23, 9 November 1989, pp. 1609-1611.

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

[5] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978.

[6] B. Yegnanarayana, "Formant extraction from linear prediction phase spectra", *J. Acoust. Soc. Amer.* Vol. 63, No. 5, May 1978, pp. 1638-1640.

[7] B. Yegnanarayana, D.K. Saikia and T.R. Krishnan, "Significance of group delay functions in signal reconstruction from spectral magnitude or phase", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-32, No. 3, June 1984, pp. 610-623.