PROCESSING OF NOISY SPEECH USING GROUP DELAY FUNCTIONS

Joy A.Thomas[+], B.Yegnanarayana, Raghuram Karinthi[++] and V.Venkateswar

Department of Computer Science and Engineering
Indian Institute of Technology, Madras-600036, India.

ABSTRACT

A new noniterative technique for all pole modelling of noisy speech is proposed. We exploit the additive and high resolution properties of a group delay function to identify the high signal to noise ratio (SNR) regions of the short time spectrum and to separate out these regions from the low SNR regions. A modified spectrum is derived from the group delay function in these regions. The new spectrum is used to derive the all pole model for the speech segment. The main advantage of this method is that it is not necessary to have prior knowledge of the noise or speech characteristics as in the other methods of processing noisy speech.

I. INTRODUCTION

The objective of this paper is to introduce a new method of determining an all pole model for noisy speech. The method exploits the properties of a group delay function [1] to identify the high signal to noise (SNR) regions of the spectrum. The spectrum is these regions is used to determine the shape of the spectrum in the low SNR regions. From the modified spectrum, an all pole model is derived using the frequency domain formulation of linear prediction (LP) analysis [2].

II. THEORY OF GROUP DELAY PROCESSING

In this section we summarize some of the basic properties of group delay functions. In this paper we will deal mainly with the group delay function derived from the magnitude spectrum which is defined as the negative derivative of the phase of the minimum phase equvalent of the signal [1]. Let $V(\omega)$ be the Fourier transform of the minimum phase equivalent of a signal.

Then

$$\ln V(\omega) = c(o) + \sum_{n=1}^{\infty} c(n) \exp(-j\omega n) \quad (1)$$

where $\{c(n)\}$ are the cepstral coefficients [3]. Writing $V(\omega) = |V(\omega)| \exp(j\theta_v(\omega))$, we get the real and imaginary parts of $\ln V(\omega)$ as

$$\ln |V(\omega)| = \sum_{n=o}^{\infty} c(n) \cos n\omega \quad (3) \quad \text{(real part)}$$

and

$$\theta_v(\omega) + 2\pi\lambda(\omega) = -\sum_{n=1}^{\infty} c(n) \sin n\omega \quad (4) \quad \text{(imaginary part)}$$

where $\lambda(\omega)$ is an integer. Therefore the group delay from magnitude is given by

$$-\theta_v'(\omega) = \tau_m(\omega) = \sum_{n=1}^{\infty} n\, c(n) \cos n\omega \quad (5)$$

Properties of the group delay function have been described in [1]. In particular there are two important properties which are relevant to our present work. The overall group delay function for a cascade of several systems is a sum of the group delay functions of the individual systems (Additive property). The group delay function corresponding to each complex pole or zero pair is significant only in the neighbourhood of the frequency of the pole or zero (High resolution property). As a result of these properties the regions of positive group delay can be used to approximate the pole part of the spectrum and the regions of negative group delay to approximate the zero part of the spectrum [4],[5].

Speech can be considered as the convolution of the vocal tract response with the excitation function. In this paper we are interested only in modelling the vocal tract response. To isolate the vocal tract information, we use the smoothed group delay function derived from the first few cepstral coefficients. The cepstral window size was chosen to be same as that of the order of the final all pole model. This corresponds to the group delay derived from the cepstrally smoothed envelope of the spectrum, which reflects the formant structure of speech. Thus the positive regions of the smoothed group delay approximate the formants and the negative regions the spectral valleys. Hence the smoothed group delay can be used to locate the high SNR Regions of

CH2118-8/85/0000-0720 $1.00 © 1985 IEEE

the spectrum.

It has been shown that the entire magnitude spectrum can be reconstructed from the unsmoothed group delay [5]. But before this reconstruction we now separate out the high SNR regions of the spectrum. It is known that a pole or zero in the z-domain affects the group delay function mainly in the neighbourhood of the pole or zero [1]. Hence to remove the effect of the pole or zero, we need only to remove the group delay in that region alone, by setting it to the average value over that region. The magnitude spectrum is then reconstructed from the modified group delay function. This modified spectrum is subjected to normal LP analysis using the autocorrelation coefficients derived from the inverse Fourier transform of the modified spectrum. Since the envelope of the modified speech usually has a greater dynamic range (peak to valley ratio) than the envelope of the original noisy spectrum, a better model of the signal is obtained from the modified spectrum than from the noisy spectrum.

III. IMPLEMENTATION OF GROUP DELAY PROCESSING

The block diagram for the various steps involved in the group delay processing is shown in Fig.1. The above procedure was applied to natural and synthetic speech segments. The speech data was sampled at 10 kHz. In all our examples we have chosen 20 msec speech segments for analysis. The synthetic data was generated by exciting a sixth order all pole filter with periodic impulses at a frequency of 125 Hz. Additive Gaussian noise was added to the clean speech to generate speech at different SNRs.

Fig.2 illustrate the stages in the group delay processing of a segment of synthetic speech. Fig.2a shows the short time spectrum of the segment and Fig.2b shows the corresponding group delay function. The group delay function is modified by setting the values of the function in the spectral valley regions to zero. The spectral valley regions are determined from the smoothed group delay function. The modified group delay function and the resulting modified spectrum are shown in Fig.2c and Fig.2a respectively. The LP spectra derived from the original spectrum and from the modified spectrum are also shown in Fig.2a. The figures show that group delay processing has little effect on the LP spectrum for clean speech.

Fig.3 illustrates the result of group delay processing for a segment of natural voiced speech corrupted by additive noise at an SNR of 10 dB. The figure shows the standard LP spectrum, the LP spectrum for noisy speech and the LP Spectrum after group delay processing. The results show that it is possible to derive an all pole model for noisy speech through group delay processing without using an iterative procedure as in [6]. The results also suggest that it is not necessary to separately extract the characteristics of noise to derive the model. Result of group delay processing of a segment of unvoiced speech corrupted by additive noise is shown in Fig.4. In both cases the LP spectrum derived after group delay processing has larger dynamic range compared to the LP spectrum derived directly from noisy speech.

IV. CONCLUSIONS

The technique proposed in this paper offers substantial improvement in the all pole modelling of noisy speech as compared to the conventional LP analysis. The improvement is most noticeable at low signal to noise ratios. Through group delay processing it is possible to derive the spectral valley regions from the information in the peak (formant) regions and this helps in retrieving to some extent the dynamic range of the original spectrum from noisy speech spectrum. Our ability to manipulate different regions of the spectrum independent of each other is a major advantage, something that cannot be done directly in the spectral domain. Also, this method does not make use of any statistical characteristics of noise to reconstruct speech from noisy data. We are presently working in the development of a complete speech analysis/synthesis system using group delay processing.

REFERENCES

[1] B.Yegnanarayana, 'Formant extraction from linear prediction phase spectra', J.Acoust.Soc.Amer., Vol. 63, pp.1638-1640, May 1978.

[2] J.Makhoul, 'Linear Prediction - A tutorial review', Proc.IEEE, Vol.63, pp.561-580, April 1975.

[3] A.V.Oppenheim and R.W.Schafer, 'Digital Signal Processing', Englewood Cliffs ,NJ, Prentice-Hall, 1975, Ch.7.

[4] 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 Processing, pp.610-623, June 1984.

[5] B.Yegnanarayana, 'Speech analysis by pole-zero decomposition of short-term spectra', Signal Processing, Vol.3, pp5-17, January1981.

[6] J.S.Lim and A.V.Oppenheim, 'All pole modelling of degraded speech', IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-26, pp.197-210, June 1978.

+ Department of Electrical Engineering, Stanford University, Stanford, CA94305, U.S.A.

++ Computer Science Department, University of Maryland, College Park, MD20742, U.S.A.

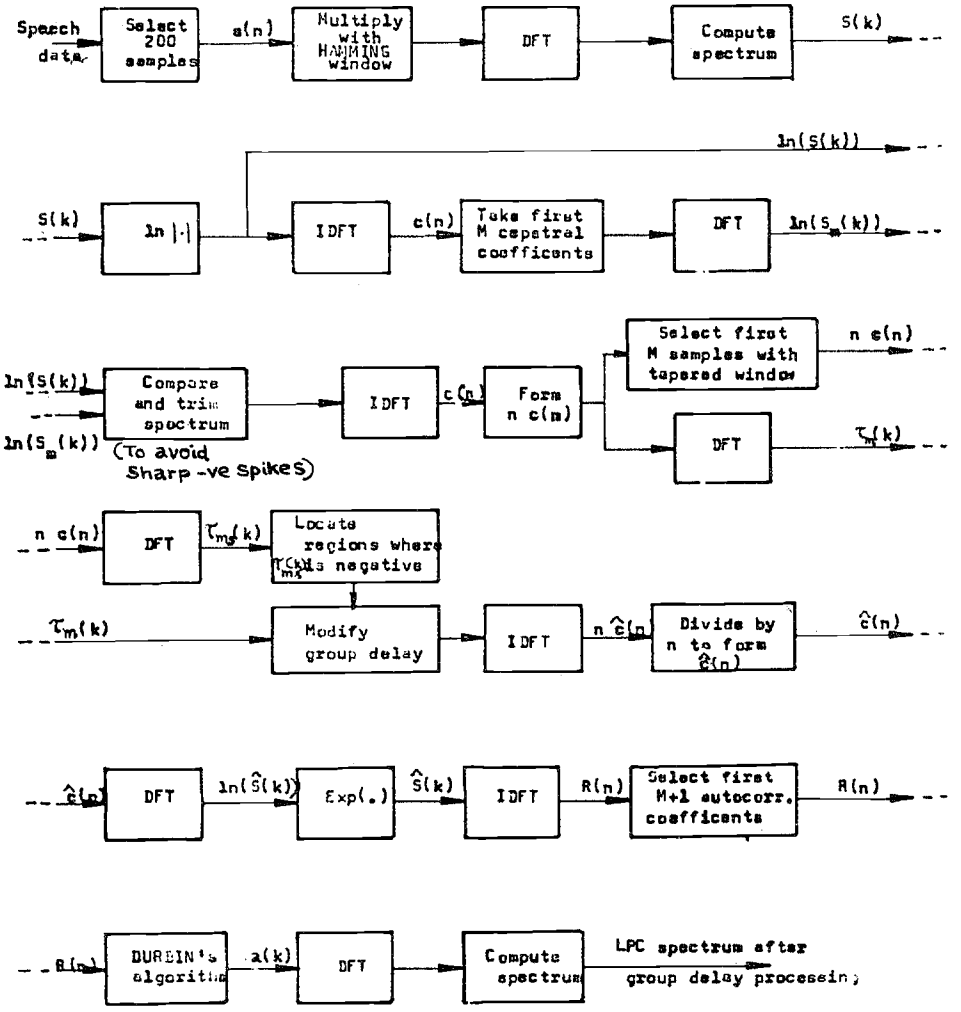

Fig.1 Block diagram showing the details of group delay processing of noisy speech

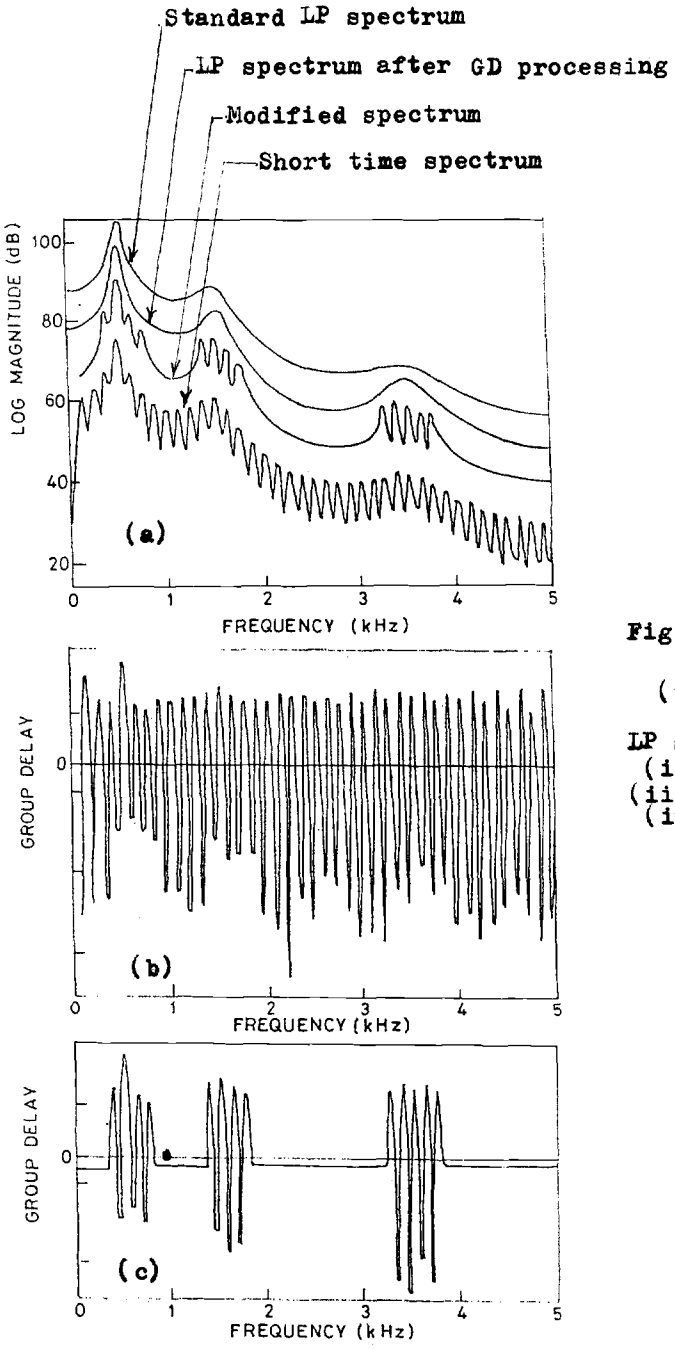

Fig.2 Various stages of group delay processing of speech.
(a) Log spectra of a short segment of speech before and after group delay processing.
(b) Group delay function corresponding to short time spectrum of speech.
(c) Modified group delay function.

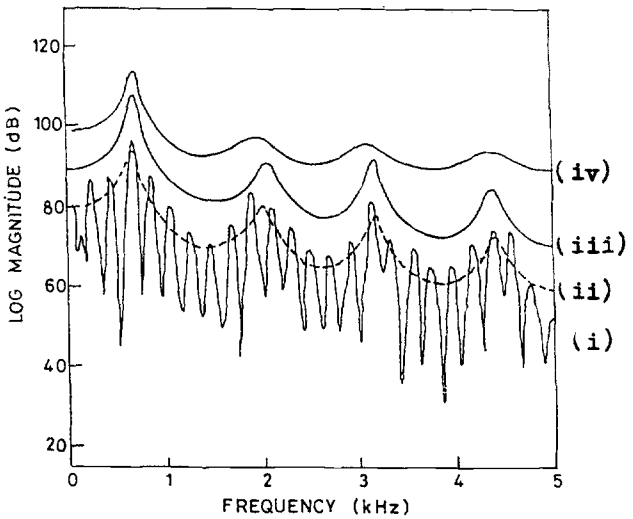

Fig.3 Group delay processing of noisy voiced speech segment at SNR = 10 dB.
(i) Log spectrum of the clean voiced speech segment.
LP spectrum of
(ii) clean speech segment
(iii) noisy segment after GD processing
(iv) noisy segment without GD processing

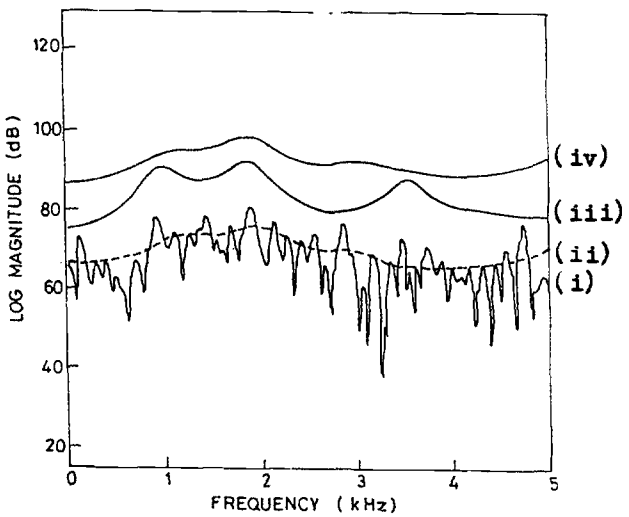

Fig.4 Group delay processing of noisy unvoiced speech segment at SNR = 10 dB
(i) Log spectrum of the clean unvoiced speech segment.
LP spectrum of
(ii) clean speech segment
(iii) noisy segment after GD processing
(iv) noisy segment without GD processing