FORMANT EXTRACTION FROM FOURIER TRANSFORM PHASE

Hema A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana

Department of Computer Science and Engineering,
Indian Institute of Technology, Madras - 600 036, India.

ABSTRACT

In this paper we propose a method of extracting formant information from short-time Fourier transform (FT) phase spectrum of speech. Fourier transform phase has not been used for formant extraction because it appears to be noisy and difficult to interpret. The effects of wrapping of phase (due to zeros close to the unit circle and the linear phase component) make it difficult to derive useful information. We develop algorithms to reduce the effects of wrapping due to linear phase and zeros close to the unit circle in the z-plane.

I. INTRODUCTION

The objective of this study is to explore the possibility of using the short-time Fourier transform phase spectrum in extracting formant information from speech data. Conventionally formants (resonance frequencies of vocal tract) are identified with peaks in the smoothed short-time FT magnitude function of a speech segment. Identification of these peaks depends on several factors such as the method used to obtain the smoothed magnitude function, the relative amplitudes and frequencies of the resonances, the method used to pick the peaks and the parameters (such as size and shape of the window) used for processing the speech signal. Difficulties in formant extraction from the smoothed magnitude function sometimes arise due to the effects of pitch periods in case of long segments and due to loss of resolution in case of short segment analysis.

In model-based techniques like linear prediction analysis there may be some spurious peaks if the model order is large. Different components, such as spectra of the individual resonances and that of the excitation, are multiplied in the FT magnitude function. This multiplicative behaviour largely dictates the performance of any formant extraction method.

In short-time Fourier analysis of real data, we have both magnitude part and phase part, although the phase part is rarely processed for information extraction. Phase function appears noise-like and difficult to process because it is usually in wrapped form (confined to the interval $\pm\pi$). But one significant feature of the phase function is that the component phase spectra are additive, unlike in the case of magnitude spectrum. While it is not clear at this time how to exploit this feature effectively, the phase function can be processed to obtain information similar to that obtained from the magnitude function.

Our recent 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[1]. In particular, for a general mixed phase signal, the contribution of zeros outside the unit circle in the z-plane have opposite signs in the group delay functions derived from magnitude and phase functions [1]. Since formants correspond to poles within the unit circle, their contribution in both the group delay functions are identical. The main problem is to extract the formant information from the data mixed with contributions from other components such as excitation, windowing, noise etc.

The motivation for our study is due to the formant features that show up in phase functions for some speech segments as shown in Fig.1. The phase function shows rapid transition around the resonant frequencies. These features are seen even for some noisy speech segments as shown in the illustration in Fig.2. However, these are carefully chosen segments to illustrate the point. In general, the phase function appears as shown in Fig.3, where the phase transitions due to formants are masked by other components. The most important and difficult issue in phase processing is the wrapping that occurs in the computation of phase from the real and imaginary parts of the FT. The extent of wrapping depends on the value of linear phase component and the location of zeros in the z-plane relative to the unit circle. The location, shape and size of the analysis window relative to the major excitation in a pitch period, besides the nature of the excitation signal, significantly influence the location of zeros of the z-tranform of the signal in the z-plane. In addition, any additive noise in the signal alters the location of the zeros.

The purpose of this paper is to suggest methods to reduce the effects of various factors responsible for wrapping of the phase function, so that the phase transitions corresponding to the formants can be extracted. Our objective is mainly

CH2673-2/89/0000-0484 $1.00 © 1989 IEEE

to demonstrate that careful anlysis of phase data provides useful information about formants.

II. COMPUTATION OF DERIVATIVE OF PHASE FOR FORMANT EXTRACTION

2.1. Group Delay Functions

For a minimum phase signal the FT magnitude and phase are related in the sense that one can be derived from the other through Hilbert transform [2]. This relation can be seen more directly through group delay functions [1]. For a minimum phase signal, the group delay function $\tau_m(\omega)$ derived from magnitude spectrum and the group delay function $\tau_p(\omega)$ derived from the phase spectrum are identical [1]. Fig.4 shows various spectra for a synthetic signal consisting of impulse response of cascade of five resonators. For this case $\tau_m(\omega)$ and $\tau_p(\omega)$ are identical. Figs.5 and 6 show the corresponding plots for windowed synthetic signals consisting of single and multiple overlapping impulse responses. The linear phase component and the distribution of zeros in the z-plane mask the phase transitions and the corresponding resonance peaks in Figs.5c and 6c respectively.

2.2. Effect of Linear Phase Component

Presence of large linear phase component makes it difficult to visualise the features of resonances in the phase due to wrapping. But in the group delay function τ_p, the linear phase component shows up as a constant. Hence the shape of τ_p can be used to interpret resonance peaks. In order to compute τ_p digitally we propose the following procedure:

Let x(n) be the time signal and $X_R(\omega)$ and $X_I(\omega)$ be the real and imaginary parts of the FT of x(n). Also let $\theta(\omega)$ and $\theta(\omega+\delta\omega)$ are the unwrapped phase function values at ω and $\omega+\delta\omega$. Then we have the following relations:

$$\tan(\tau_p(\omega)\delta\omega) = \tan[\theta(\omega+\delta\omega) - \theta(\omega)]$$

$$= \frac{\tan\theta(\omega+\delta\omega) - \tan\theta(\omega)}{1 + \tan\theta(\omega+\delta\omega)\tan\theta(\omega)} \quad (1)$$

where

$$\tan\theta(\omega) = X_I(\omega) / X_R(\omega) \quad (2)$$

$$\tan\theta(\omega+\delta\omega) = X_I(\omega+\delta\omega) / X_R(\omega+\delta\omega) \quad (3)$$

$$X_R(\omega) + j X_I(\omega) = \sum_n x(n)\, e^{-j\omega n} \quad (4)$$

$$X_R(\omega+\delta\omega) + j X_I(\omega+\delta\omega) = \sum_n x(n)\, e^{-j(\omega+\delta\omega)n} \quad (5)$$

Since $\delta\omega$ is constant, it is possible to obtain $\tau_p(\omega)$ which reflects the resonance characteristics. Fig.5d shows the processed $\tau_p(\omega)$ for the signal shown in Fig.5a.

2.3. Effect of Zeros Close to the Unit Circle

For most practical signals, there are a large number of zeros close to the unit circle in the z-plane. They produce large fluctuations in τ_p, which will make interpretation of this function difficult for resonance peak estimation. One possible solution is to implicitly smooth the phase spectrum and then compute the group delay function. This can be done by taking the average of complex spectral values at (say) three points $\omega-\delta\omega$, ω and $\omega+\delta\omega$ and then computing the phase at ω. This averaging can be done equivalently in the time domain by generating the signal

$$\tilde{x}(n) = \frac{1}{3}\left[x(n)\, e^{jn\delta\theta} + x(n) + x(n)\, e^{-jn\delta\theta} \right] \quad (6)$$

where $\delta\theta$ is an angular shift.

The steps in the computation of the group delay function $\tau_p(\omega)$ are as follows:

i. Compute $\tilde{x}(n)$. Choose $\delta\theta$ as a fraction of the frequency sampling interval.

ii. Compute the DFTs $\tilde{X}(k)$ and $\tilde{X}(k+\delta\omega)$ of $\tilde{x}(n)$ and $\tilde{x}(n)\, e^{-jn\delta\omega}$ respectively, where $\delta\omega \leq 2\pi/N$.

iii. Compute the group delay function $\tau_p(k)$ using the real and imaginary parts of $\tilde{X}(k)$ and $\tilde{X}(k+\delta\omega)$.

iv. Perform cepstral smoothing on $\tau_p(k)$ by computing IDFT of $\tau_p(k)$, truncating the sequence, and recomputing the DFT of the truncated sequence.

v. Find the peaks in the resulting smoothed group delay function corresponding to formants.

Fig.6d shows the smoothed group delay obtained for the example in Fig.6a.

III. ILLUSTRATION WITH SPEECH DATA

Fig.7 shows the comparison of the smoothed group delay function $\tau_p(\omega)$ with the linear prediction spectrum as well as the standard cepstrally smoothed spectrum. This example illustrates that the phase function can be processed to obtain useful information. However, there are several cases where zeros outside the unit circle in the z-plane will produce spurious peaks and also mask the true formant peaks.

IV. CONCLUSIONS

In this paper we have shown that formant information can be obtained from smoothed group delay function derived from FT phase. The studies have demonstrated that the FT phase is as important as the FT magnitude function. If proper techniques for phase unwrapping or for obtaining group delay are available, it may be possible to exploit the additive property of phase functions for many signal processing applications. However, reliable and consistent estimation of formant data requires optimization of several analysis parameters such as window function, frequency spacing, etc., besides sophisticated techniques for dealing with zeros outside the unit circle in the z-plane.

REFERENCES

1. B.Yegnanarayana, D.K.Saikia, and T.R.Krishnan, 'Significance of Group Delay Functions in Signal Reconstruction from Spectral Magnitude or Phase,' IEEE Trans. Acoustics, Speech and Signal Proc., Vol.ASSP-32, No.3, pp.610-623, June 1984.

2. A.V.Oppenheim, R.W.Schafer, Digital Signal Processing, Englewood Cliffs, NJ, Prentice-Hall, 1975.

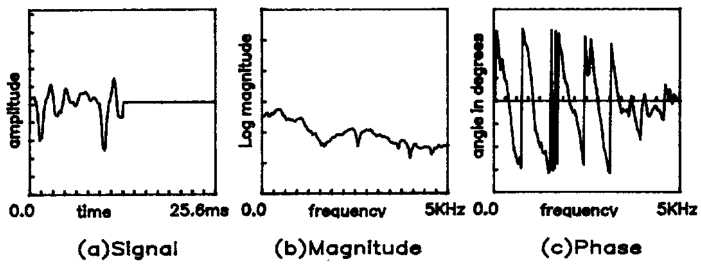

Fig.1 Illustration of phase transitions for a selected segment of speech.

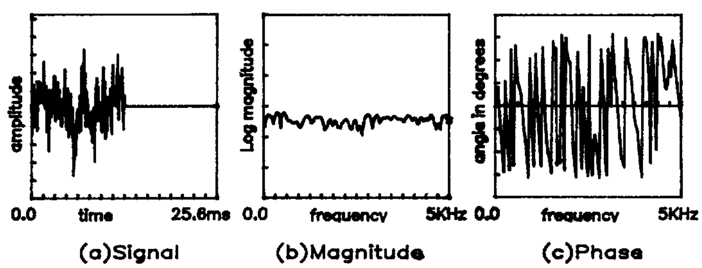

Fig.2 Illustration of phase transitions for a selected segment of noisy speech.

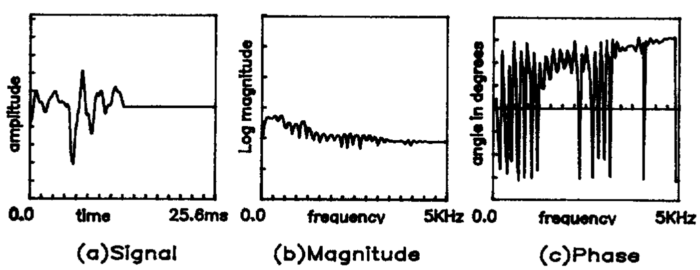

Fig.3 Illustration of phase transitions for an arbitrary segment of speech.

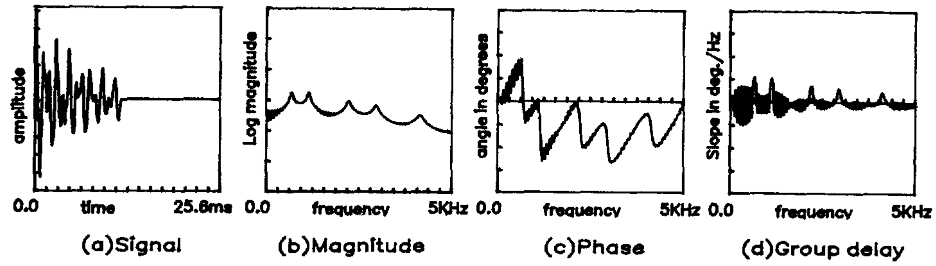

Fig.4 Illustration of the relation between log-magnitude, phase and group delay spectra for an impulse response of a five resonator vocal tract model.

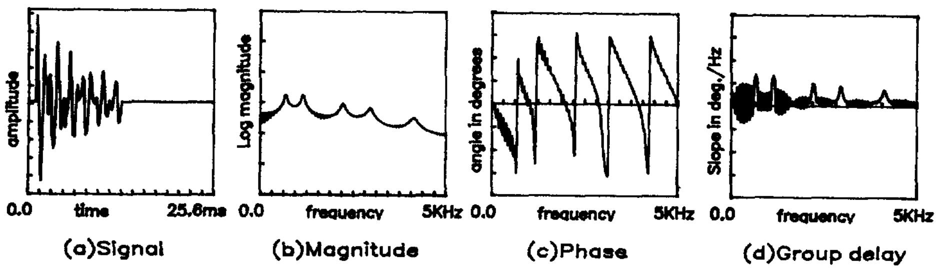

Fig.5 Illustration of the effect of linear phase on phase transitions for a synthetic signal. Fig.5d shows the processed group delay.

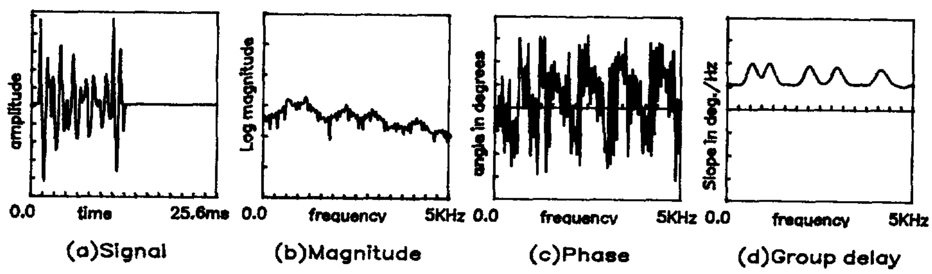

Fig.6 Illustration of the effect of zeros close to the unit circle on phase transitions for a synthetic signal. Fig.6d gives the processed group delay.

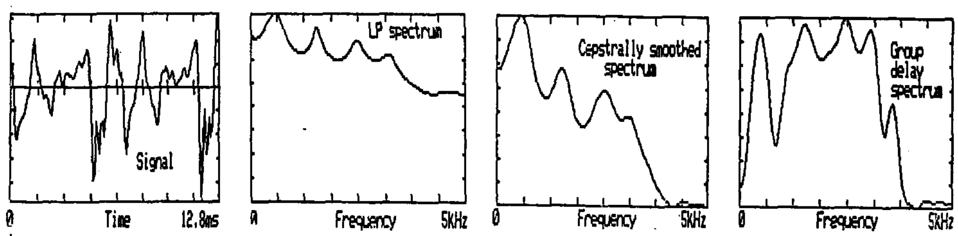

Fig.7 A segment of speech and its smoothed spectra obtained using three different techniques.