ALGORITHMS FOR PROCESSING FOURIER TRANSFORM PHASE OF SIGNALS

A ***THESIS***

Submitted for the award of the Degree

of

DOCTQR QF PHILQSQPHY

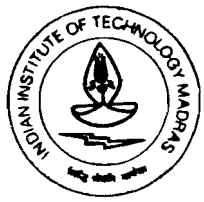
in

COMPUTER SCIENCE AND ENGINEERING

by

HEMA A. MURTHY

SPEECH AND VISION LABORATORY
CSE, IIT, Madras - 600 036



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
MADRAS-600 036, INDIA

DECEMBER 1991

ALGORITHMS FOR PROCESSING FOURIER TRANSFORM PHASE OF SIGNALS

CERTIFICATE

This is to certify that the thesis entitled "ALGORITHMS FOR PROCESSING FOURIER TRANSFORM PHASE OF SIGNALS" is the bonafide work of Hema A. Murthy, carried out under my guidance and supervision, at the Department of Computer Science and Engineering, Indian Institute of Technology, Madras, for the award of the degree of Doctor of Philosophy in Computer Science and Engineering.

(B. YEGNANARAYANA)

ABSTRACT

The studies presented in this thesis represent an attempt to process the Fourier transform (FT) phase of signals for feature extraction. Although the FT magnitude and phase spectra are independent functions of frequency features of a signal, most techniques for feature extraction from a signal are based upon manipulating the the FT magnitude only.

The phase spectrum of the signal corresponds to time delay corresponding to each of the sinusoidal components of the signal. In the context of additive noise, the time delay may not be significantly corrupted and the phase spectrum might be considered to be a more reliable source for estimating the features in a noisy signal. Although the importance of phase in signals is realised by researchers, very few attempts have been made to ***process*** the FT phase of signals for the extraction of features. Features of a signal, for example, resonance information, is completely masked by the inevitable wrapping of the phase spectrum.

An alternative to processing the phase spectrum is processing the group delay function. The group delay function is the negative derivative of the (unwrapped) FT phase spectrum. The group delay function can be computed directly from the time domain signal. The group delay function possesses ***additive*** and ***high resolution*** properties, in that it shows a squared magnitude behaviour in the vicinity of a resonance. But the group delay function in general is not well behaved for all classes of signals. Zeros in the z-transform of a signal that are close to the unit circle cause large amplitude spikes to appear in the group delay function. The polarity of a spike depends on the location of the zero with respect to the unit circle. These large amplitude spikes mask the information about

resonances.

The research effort in this thesis focusses on the development of algorithms for manipulating the group delay function to suppress the information corresponding to the zeros of th signal that are close to unit circle in the z-domain and emphasise the features of of a signal. To demonstrate the usefulness of the algorithms developed, these algorithms are used to estimate (a) formant and pitch data from speech signals and (b) estimate spectra of auto-regressive processes and sinusoids in noise.

The research effort in this thesis shows that the phase spectrum (or rather the group delay function) of a signal can be usefully processed to reliably extract features of a signal.

ACKNOWLEDGEMENT

I express my appreciation to Prof. B. Yegnanarayana for his constant help, excellent guidance and constructive criticisms throughout the course of this work. I thank Prof. R. Nagarajan, Head, Department of Computer Science and Engineering, for making the various facilities in the department available to me.

I owe my special thanks to Madhu Murthy and C. P. Mariadassou for some fruitful discussions. I thank G. V. Ramana Rao and R. Ramaseshan for reading my thesis and making useful suggestions. I would like to thank all my colleagues of the Speech and Vision Lab who have helped me in one way or the other.

I thank Vatsala for providing me a shoulder whenever I was depressed.

Finally, I thank my husband M. V. N. Murthy for his support and perseverence throughout the course of this work.

CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

OVERVIEW OF THE THESIS

1.1 *Introduction*

This thesis represents an attempt to extract features of a signal by processing the Fourier transform (FT) phase spectrum of a signal rather than the conventional FT magnitude spectrum of a signal. As a result, algorithms based on manipulating the FT phase are developed and applied in speech analysis and spectrum estimation.

The Fourier representation of a signal is complete only when both the spectral magnitude and phase are specified. However, under certain conditions, the signal can be completely specified by the FT magnitude (to within a time shift) or by the FT phase (to within a scale factor). Information, such as resonance characteristics of a signl is present both in the FT magnitude and FT phase. But most techniques for estimating the parameters of a signal, especially, parameters of the source and system in a source-system model for signal production are based upon processing the FT magnitude spectrum only.

The phase spectrum of the signal corresponds to time delay corresponding to each of the sinusoidal components of the signal. In the context of additive noise, the time delay may not be significantly corrupted and the phase spectrum might be considered to be a more reliable source for estimating the features in a noisy signal. In multidimensional signal processing it has been shown that features about the signal like edges are preserved better in the phase spectrum than that of the magnitude spectrum.

Although the importance of phase in signals is realised by researchers, very few attempts have been made to ***process*** the FT phase of signals for the extraction of features. The reason for analysis

techniques being based on processing FT magnitude rather than FT phase is that it is possible to visually perceive the features in a signal in the magnitude spectrum. For example, the resonances in a signal manifest as peaks of envelope of the magnitude spectrum, while they manifest as transitions of phase in the phase spectrum. The peaks in the envelope of the magnitude spectrum are visible while the phase transitions are completely masked by the inevitable wrapping of the phase spectrum.

Therefore, an alternative to processing the phase spectrum is processing the group delay function. The group delay function is the negative derivative of the (unwrapped) FT phase spectrum. The group delay function can be computed directly from the time domain signal. For a minimum phase signal the group delay function shows a squared magnitude behaviour around a resonance/antiresonance frequency. In addition, the information about a resonance/antiresonance is concentrated around the resonant/antiresonant frequency. These properties of the group delay function are referred to as the additive and high resolution properties. But the group delay function in general is not well behaved for many signals. Zeros in the z-transform of a signal that are close to the unit circle cause large amplitude spikes to appear in the group delay function. The polarity of a spike depends on the location of the zero with respect to the unit circle. These large amplitude spikes mask the information about resonances.

Speech signal can be modelled as the response of a time varying minimum phase digital filter (generally all-pole) to an impulse or random noise excitation. The group delay function of a segment of speech has very large amplitude spikes that are caused by the zeros due to the excitation and the finite duration of the segment (data

window). In the context of spectrum estimation, the group delay function of an Auto-regressive (AR) process in noise also has very large amplitude spikes that are due to the data window and random noise excitation. Therefore, if the resonance behaviour of a signal is to be studied through group delay functions, the spikes caused by the zeros must be eliminated. This suggests that the focus of algorithms for processing group delay functions of signals should be (a) the removal of the effects of zeros due to excitation and (or) data window and (b) separation of system and source components in the group delay domain.

1.2 Scope of *the thesis*

The research effort in this thesis focusses on the development of algorithms for processing the processing the FT phase through group delay function, in a systematic manner. Because of the manner in which the different components of a signal combine in the FT phase, the results obtained are not necessarily identical to those obtained by processing the FT magnitude spectrum. In some cases, the information obtained may reinforce the information obtained from the magnitude spectrum. To demonstrate the usefulness of the algorithms developed, examples are chosen from (a) analysis of speech (synthetic and natural) signals and (b) spectrum estimation.

If it can be assumed that the system in a source-system model of signal production is minimum phase, it should be possible to obtain an estimate of the system characteristics by estimating the minimum phase component of the signal. Linear prediction (LP) analysis is an approach for processing signals, where an attempt is made to estimate the parameters of a the model of the minimum phase system from the signal. In LP analysis the system is asssumed to be an all-pole system while the source may be a train of impulses or random noise.

The resonance information is then obtained from the estimated parameters of the system. Adopting a similar strategy, a signal with minimum phase characteristics is derived from the short-time Fourier transform (STFT) magnitude spectrum of the signal. The group delay function of this signal is then computed in which the information about the resonances of the system is preserved. This group delay function is henceforth referred to as the minimum-phase group delay function. This algorithm is then used to extract formants from speech. To demonstrate the effectiveness of this technique for formant extraction from speech signal the following studies are made:

(a) Performance of the formant extraction based on the minimum phase group delay function is evaluated for a synthetic signal. The synthetic signal is obtained using a formant vocoder in which the formants are continuously changed to reflect the formant transitions that occur in natural speech.

(b) Comparison of the minimum phase-based group delay method with standard linear prediction (LP) analysis and cepstrum analysis.

It was observed that the formant extraction based on the minimum phase group delay :

(i) tracks formant changes well.

(ii) gives more consistent estimates of formants when compared with that of LP analysis and cepstrum Analysis for various choices of analysis parameters.

The technique just described still uses the group delay function of a minimum phase signal which is in turn derived from the magnitude spectrum. This is because the group delay function suffers from poor sampling when the zeros of the signal z-tranform have zeros that are close to the unit circle (both inside and outside the unit circle) in the z-domain. Ideally it would be desirable to separate the minimum

and nonmimum phase components of a signal in some domain inorder to estimate parameters corresponding to that of the minimum and nonminimum phase components of a signal.

Cepstrum analysis is a method in which it is possible to approximately separate the system and source components of a mixed phase signal. In this method the system and source components that are multiplicative in the FT magnitude spectrum become additive in the cepstral domain. In addition, the source and system components are well separated in cepstrum. This enables the use of a gating function in the cepstral domain to separate the system and source information.

Because of the nature of the FT phase spectrum, the system and source components are additive in the group delay function. But the system and source information is spread over the entire function. Therefore it may not be possible to separate them at all in the group delay domain.

Although it may not be possible to completely separate the two components of a signal, it may be possible to suppress one while favouring the other. This is exactly what is done in the modified group delay (MGD) function that is derived from the group delay function of the signal. To estimate the components corresponding to that of the system, a modified group delay function is obtained in which the group delay information corresponding to the source is suppressed. To estimate the components corresponding to that of the source, another group delay function is derived in which, the information corresponding to that of the system is suppressed. Properties of the MGD are studied for synthetic and natural speech signals. Both MGDs are then used to estimate formants and pitch from speech.

From another viewpoint the MGD may be thought of as an function in which the zeros of a signal are suppressed. Additive Gaussian noise either introduces new zeros or redistributes existing zeros of the signal. If the noise level is not too high, it should be possible to estimate the system and source parameters using the MGD. Formant and pitch data are estimated using the MGD from noisy speech. The formant and pitch data can then be used to synthesise speech.

Application of the MGD for problems in spectrum estimation is studied. In particular, in spectrum estimation, two different examples are considered, namely. (a) sinusoids in noise and (b) autoregressive process in noise. Using the relationship between the cepstrum and the group delay function, the power spectrum corresponding to the system is derived. It is observed.in the power spectra derived for sinusoids both bias and variance are significantly reduced in the estimated power spectrum throught group delay function compared to that of the periodogram estimates of the power spectrum. For the AR process, bias and sidelobe leakage are reduced. But the variance around the resonances is higher than that of the periodogram approach.

The research effort in this thesis shows that the phase spectrum of a signal can be usefully processed to realiably extract features of a signal.

1.3 *Organisation of the thesis*

The thesis is organised as follows. In Chapter 2 we briefly discuss the time and frequency domain representations of signals giving special emphasis to the FT. We bring out the relationship between FT magnitude spectrum and FT phase spectrum of a signal through group delay functions. We discuss the properties of the group delay function and the problems of processing signals like

speech using group delay functions. Digital representation of speech signals is discussed. Issues in speech analysis and spectrum estimation for feature extraction are also discussed.

We develop a new algorithm for formant extraction from speech using a minimum phase-based group delay function derived from the STFT magnitude spectrum in Chapter 3.

In Chapter 4 we derive a modified group delay function directly from the signal. Properties of this modified group delay are studied in detail. Algorithms (based on the modified group delay function) are developed for formant and pitch extraction from speech. Formant and pitch data are also extracted from noisy speech. Application of the MGD in spectrum estimation is studied in Chapter 5.

Finally in Chapter 6 we summarise the results of the investigations done in this thesis. A brief discussion of the major contributions and drawbacks of this thesis are also given in this Chapter.

Some of the relevant derivations used in this thesis are derived in the Appendices.

1.4 *Major Contributions* of *the thesis*

The following are the major contributions of the thesis:

(a) A new algorithm for formant extraction from speech using a group delay function derived from the FT magnitude spectrum.

(b) A new algorithm for formant extraction from speech using a modified group delay function derived from FT phase.

(c) A new algorithm for pitch extraction from speech using the modified group delay function.

(d) A new algorithm for spectrum estimation using modified group delay function.

CHAPTER 2

REPRESENTATIONS OF SIGNALS

2.1 Introduction

Signals are basically quantities that fluctuate with time. It is natural and convenient for us to think of signals as functions of time. An optical image on the other hand may be described by a function of spatial coordinates. A familiar representation, in fact one that is deeply ingrained by usage, is the graph of a function. The graph is a collection of ordered pairs of numbers {t,x(t)}. From the standpoint of system design, the graphical representation is unmanageable simply because it consists of too many individual points.

In contrast to a graphical representation, where signals are represented by a collection of points in a simple setting a two-dimensional space, a more highly structured setting is the signal space, whereby a signal can be considered as a single entity or a point in a space. For example, consider, the representation of a sinusoid in terms of its frequency components. Representation of a sinusoid in the time domain requires an infinite number of points. In the frequency domain it can be represented by a single point.

The primary objective of this Chapter is to examine some of the issues involved in digital processing of signals using group delay functions. To achieve this in Section 2.2 an introduction to the analysis of signals in terms of the continuous Fourier transform is discussed. Although the continuous Fourier transform is useful in giving an interpretation to a signal it is not suitable for practical applications.

To enable the analysis of signals by computer, the signals are sampled and quantised. The signals are represented by a sequence of

numbers which are obtained by sampling the analog signal at discrete time intervals. This has led to a new field of study "Digital Signal Processing". Digital signals are analysed by the computer using a discrete Fourier transform (or DFT) and inverse DFT. The conditions under which the discrete time signal and the DFT are exact representations of the continuous time signal and continuous FT are also discussed in Section 2.2.

The Fourier transform is extensively used in signal analysis as a tool to resolve a given signal into its sinusoidal/complex exponential components. Fourier transform of a signal is in general complex. In polar form it consists of two parts, a magnitude part and a phase part which are called Fourier transform magnitude spectrum and Fourier transform phase spectrum respectively. The FT magnitude and phase spectrum are in general distinct functions of frequency. Properties of the FT phase and magnitude spectrum that are relevant in this study are listed in Section 2.3.

Relationship between the FT magnitude and FT phase is brought out in Section 2.4 through group delay functions. The group delay functions have some useful properties, that can be exploited in signal analysis. These properties are studied in Section 2.4. Although the usefulness of the group delay representation of signals has been established in the literature, the standard group delay function is not suitable for the representation of natural signals like speech. The background for speech processing is given in Sections 2.5 and 2.6. The problem of spectrum estimation when no apriori information is available about the signal is addressed in Section 2.7. The motivation for the work done in this thesis is presented in Section 2.8. Issues involved in the application of group delay functions to speech processing and spectrum estimation

are addressed in Section 2.9.

2.2 *Fourier representation of Signals*

2.2.1 Significance of the Fourier transform

Fourier transforms play an important part in the theory of many branches of science. Mathematically the Fourier transform is a functional, i.e. it is a mapping from an arbitrary set of functions into another set of functions. The physical meaning of the Fourier transform is that it enables analysis of time functions in terms of their spectra or frequency content. It basically enables (a) the analysis of a signal in terms of its various frequency components and (b) the synthesis of a signals from its sinusoidal components. A waveform optical, electrical, or acoustical and its spectrum are appreciated equally as physically picturable and measurable entities. For example, in the theory of speech production, the vocal tract is characterised by a cascade of resonators. The frequencies and bandwidths of these resonators change continuously with time. This leads to the articulation of different sounds.

Estimation of the power spectral density or simply the spectrum of discretely sampled deterministic and stochastic processes is usually based on procedures employing the Fourier transform. The objective of spectrum estimation is to answer specific questions about the data. It is found [S.M. Kay, 1988, S.L. Marple; 19871 that in most applications, the frequency distribution of the signal is of interest. For example, the presence or absence of a sinusoid in a signal can be determined by looking at the frequency distribution of the signal. If the spectrum peaks at a particular frequency, it can be concluded that this sinusoidal component is present in the signal.

2.2.2. Continuous Fourier transform

The Fourier transform is a general form of a relation between

the elements of two different sets in that it is a mapping of elements from one set into elements of another set. A mapping is simply the rule by which elements of one set say S_1 are assigned to elements of the other set say S_2, Symbolically this may be denoted by

$$f : S_1 \rightarrow S_2$$

which is a compact notation for

$$y = f(x),\ x \in S_1 \quad \text{and} \quad y \in S_2$$

The element y is called the image of x under the mapping f. The set S_1 is the ***domain*** of the mapping and the set of all images of the elements of S_1 (contained in S_2) is the ***range*** of mapping.

If S_1 is the set of bounded energy signals

$$S_1 = \{ x : \int_a^b x^2(t)dt < m \}$$

then the Fourier transform $\mathcal{F} : S_1 \rightarrow S_2$ is a mapping into another set of square integrable functions

$$S_2 = \{ X : \int_{\infty}^{\infty} X^2(\omega)d\omega < \infty \}$$

This mapping is described by

$$\mathcal{F} : S_1 \rightarrow S_2 \Rightarrow X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt$$

Strictly speaking this is not a one-to-one mapping but a many-to-one mapping.

Similarly, the Inverse Fourier Transform (IFT) is defined by the following mapping :

$$\mathcal{F}^{-1} : S_2 \rightarrow S_1 \Rightarrow x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}d\omega$$

Waveforms and spectra are transforms of each other. The Fourier transform resolves a signal into its complex exponential components. The inverse Fourier transform synthesises the signal from its

exponential components.

The Fourier transform is a complex function and may be expressed as

$$X(\omega) = |X(\omega)|e^{j\theta(\omega)}$$

where $|X(\omega)|$ is the amplitude or magnitude spectrum and $\theta(\omega)$ is the phase spectrum. The FT $X(\omega)$ of $x(t)$ represents the relative amplitudes of various frequency components of $x(t)$ at ,different frequencies w.

2.2.3 Discrete Fourier transform

The Fourier transform defined in the previous section is of great theoretical importance. It is not directly suitable for practical applications, however. Continuous time and frequency variables are not compatible with the discrete nature of digital processing.

Given the importance of the Fourier transform in signal processing, a more practical way to express it is the ***discrete Fourier transform*** (DFT). We first define the z-transform of a discrete time signal which is later used to define the DFT of a signal.

The ***z-transform*** representation of a sampled signal $x(n)$ (discrete in time) is defined by the pair of equations

$$\mathcal{Z} : S_1 \rightarrow S_2 \Rightarrow \quad X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} = \frac{N(z)}{D(z)} \quad (2.1a)$$

$$\mathcal{Z}^{-1} : S_2 \quad S_1 \Rightarrow x(n) = \frac{1}{2\pi} \oint_C X(z)z^{n-1}dz \quad (2.1b)$$

where $X(z)$ is in general an infinite power series in the variable z^{-1}. The values $x(n)$ play the role of coefficients in the power series.

When the z-transform is represented as a ratio of two

polynomials N(z) and D(z) (Eq.(2.1a)), the roots of N(z) are said to correspond to the zeros of the signal while the roots of D(z) are said to correspond to the poles of the signal x(n).

The Fourier transform of a discrete-time signal (sampled signal) is given by the equations

$$X(e^{j\omega}) = \sum_{n=-\infty}^{w} x(n)e^{-jwn} \qquad (2.2a)$$

$$x(n) = \frac{}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})e^{jwn}d\omega \qquad (2.2b)$$

These equations are a special case of Eqs. (2.1). These equations are obtained by restricting the z-transform to the unit circle of the z-plane, i.e., by setting $z=e^{j\omega}$. As indicated in Fig.2.1 the frequency variable, ω, also has the interpretation as angle in the z-plane.

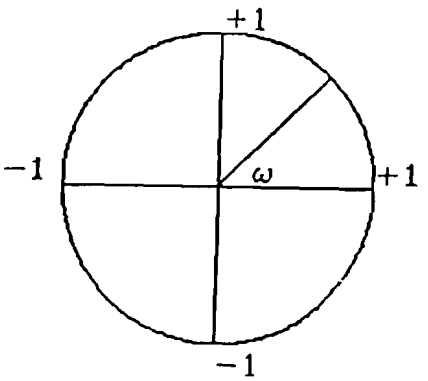


Fig.2.1 Unit circle in the z-plane

If a sequence is periodic with period N; i.e.,

$$\tilde{x}(n) = \tilde{x}(n + N) \quad -\infty < n < \infty$$

then $\tilde{x}(n)$ can be represented by a discrete sum of complex sinusoids rather than an integral equation as in Eq (2.2b). The Fourier series representation for a periodic sequence is [M.Kunt, Ch.3, 1987]

$$\tilde{X}(k) = \sum_{n=0}^{N-1} \tilde{x}(n)e^{-j\frac{2\pi kn}{N}} \qquad (2.3a)$$

$$\tilde{x}(n) = \sum_{k=0}^{N-1} \tilde{X}(k)e^{j\frac{2\pi kn}{N}} \cdot \frac{1}{N} \qquad (2.3b)$$

This is an exact representation of a periodic sequence. However, the

utility of this representation lies in imposing a different interpretation upon the above equations. Let us consider a finite length sequence x(n), that is zero outside the interval $0 \le n \le N-1$. Then the z-transform is

$$X(z) = \sum_{n=0}^{N-1} x(n) z^{-n} \qquad (2.4)$$

If we evaluate the X(z) at N equally spaced points on the unit circle, i.e., $z_k = e^{j\frac{2\pi k}{N}}$, $k = 0,1,\ldots,N-1$, then we obtain

$$\tilde{X}(e^{j\frac{2\pi k}{N}}) = \sum_{n=0}^{N-1} \tilde{x}(n) e^{-j\frac{2\pi kn}{N}}, \quad k = 0,1,\ldots,N-1 \qquad (2.5)$$

If we construct a periodic sequence as an infinite sequence of replicas of x(n),

$$\tilde{x}(n) = \sum_{r=-\infty}^{\infty} x(n + rN) \qquad (2.6)$$

then, the samples $\tilde{X}(e^{j\frac{2\pi k}{N}})$ are easily seen to be the Fourier coefficients of the periodic sequence $\tilde{x}(n)$ in Eq.(2.3). Thus a sequence of length of N can be exactly represented by a discrete Fourier transform (DFT) representation of the form

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi kn}{N}} \qquad (2.7a)$$

$$x(n) = \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi kn}{N}} \cdot \frac{1}{N} \qquad (2.7b)$$

The only difference between Eqs.(2.3) and (2.7) is a slight modification to the notation (removing the ~ symbols which indicates periodicity) and the explicit restriction to the finite intervals $0 \le k \le N-1$ and $0 \le n \le N-1$. It is important to bear in mind when using the DFT representation that all sequences behave as if they were periodic.

To use digital processing methods on analog signals such as speech. it is necessary to represent the signal as a sequence of

numbers. This is commonly done by sampling the analog signal denoted by $x_a(t)$, periodically to produce the sequence

$$x(n) = x_a(nT) \qquad -\infty < n < \infty \qquad (2.8)$$

where n takes on only integer values and T is the sampling period in seconds.

The conditions under which the sequence of samples of Eq.(2.8) is a unique representation of the original analog signal are well known and are often summarised as follows :

The *Sampling Theorem* : If a signal $x_a(t)$ has a bandlimited Fourier transform $X_a(e^{j\omega}) = 0$ for $\omega \geq 2\pi F_N$, then $x_a(t)$ can be uniquely reconstructed from equally spaced samples $x_a(nT)$, $-\infty < n < \infty$, if $T \leq 1/2F_N$. F_N is called the Nyquist Frequency.

Similarly, if $X_a(e^{j\omega})$ is to be obtained from the samples of its Fourier transform (obtained by sampling the continuous Fourier transform at equally spaced intervals in the z-plane), the signal should be time limited. The DFT obtained in the Eq.(2.7a) is periodic with a period of 2π. As mentioned earlier, the definition of the DFT requires that the time domain signal be of finite length. In many filtering and spectral analysis applications, the signals do not inherently have finite length. This inconsistency between the finite length requirement of the DFT and the reality of indefinitely long signals can be accomodated exactly or approximately through the concepts of windowing, block processing and the computation of the time dependent Fourier transform [A.V.Oppenheim and R.W.Schafer, Ch.11, 19891.

The basic steps in applying the DFT to continuous time signals are indicated in Fig.2.2. The anti-aliasing filter is incorporated to minimize the effect of aliasing when the continuous time signal is converteu to a sequence. The need for the window w(n) in Fig.2.2 is

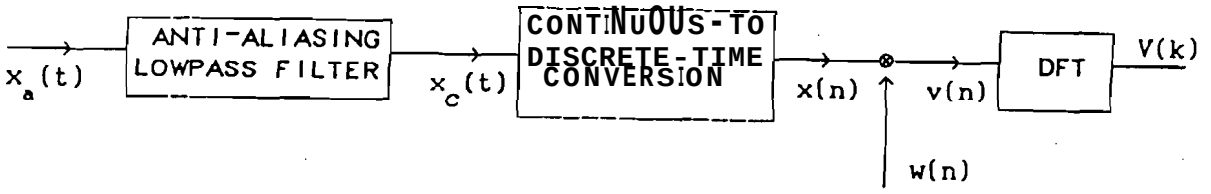


Fig.2.2(a) Processing steps in the discrete-time Fourier analysis of a continuous signal.

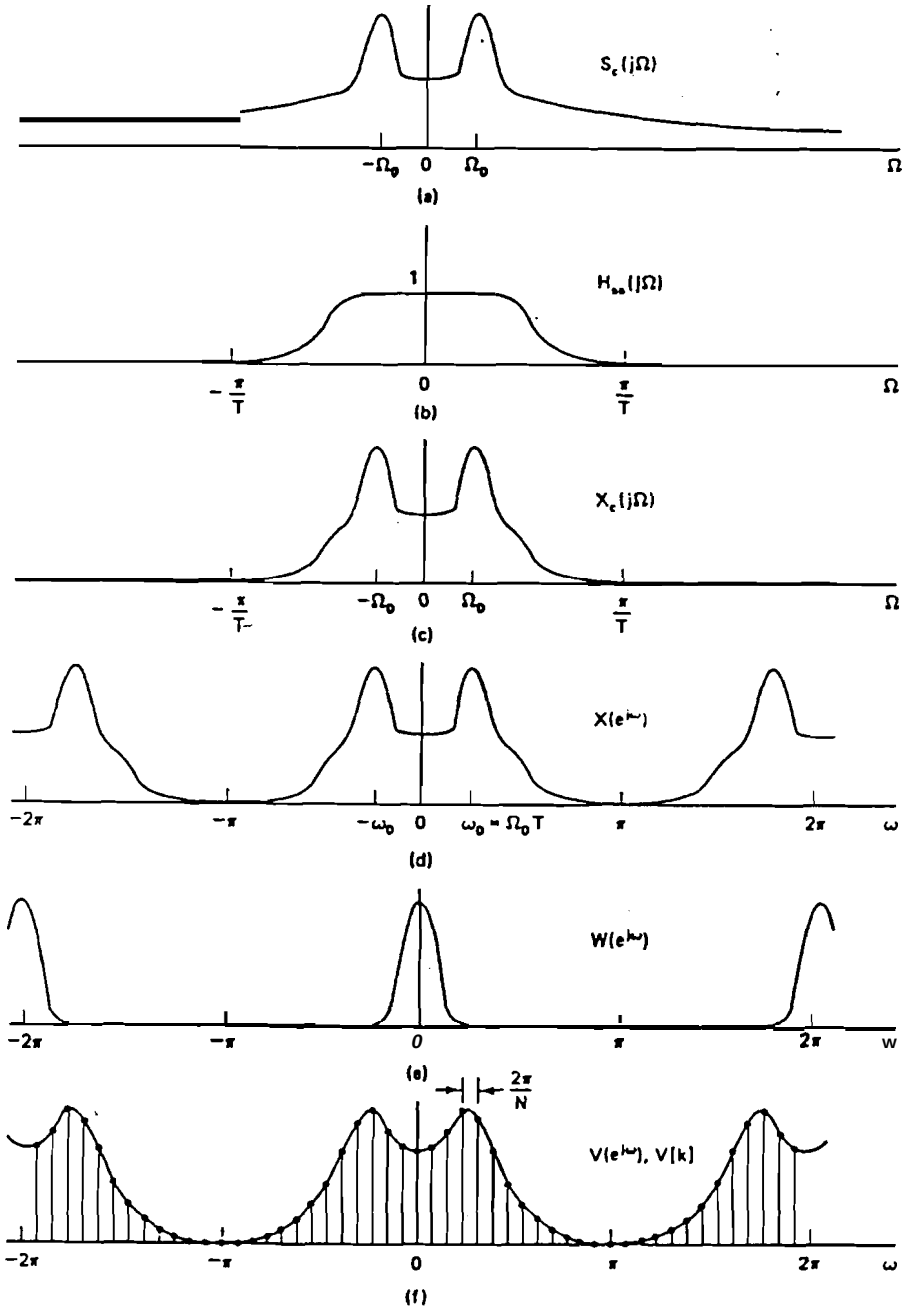


Fig.2.2(b) Illustration of the Fourier transforms in the system of Fig.2.2(a).(reproduced from [A. V. Oppenheim and R. W. Schafer; 19891) (a) Fourier transform of continuous-time input signal. (b) Frequency response of anti-aliasing filter. (c) Fourier transform of output of anti-aliasing filter. (d) Fourier transform of sampled signal. (e) Fourier transform of window sequence. (f) Fourier transform of windowed segment and frequency samples obtained using DFT samples.

a consequence of the finite length requirement of the DFT.

In the discussion that follows in the rest of this Chapter we assume that (a) sampling in the time and frequency domains is done at sufficiently close intervals to avoid aliasing and (b) the DFT and IDFT are close approximations to the continuous FT and IFT. Henceforth we use $X(\omega)$ and $X(e^{j\omega})$ interchangeably to represent the discrete time Fourier transform of the discrete time signal (sampled) x(n)), and X(k) to represent the Discrete Fourier transform of x(n).

In the rest of this chapter we restrict our discussion to the discrete time Fourier transform magnitude and phase spectra of signals.

2.3 *Properties of The Fourier transform Magnitude and Phase Spectra*

The Fourier transform Magnitude and Phase spectra are independent functions of frequency and have some distinct properties. The properties of the magnitude and the phase spectrum that are relevant in this study are listed below :

2.3.1 *Properties of the Fourier transform Magnitude Spectrum* (FTMS)

1. For any real x(n) FTMS is an even function of ω.
2. For any x(n) the FTMS is a positive function of w.
3. The Inverse Fourier Transform (IFT) of the FTMS is a noncausal even function of time. This function can be expressed as the autocorrelation function of some sequence y(n) [A. Papoulis;1977,Ch.7]. This signal is also called a zero ***phase signal.***
4. If a signal x(n) is the impulse response of a cascade of resonators and antiresonators, the overall FTMS of x(n) is the multiplication of the magnitude spectra of the individual resonators and antiresonators. The resonances are characterised by peaks in the magnitude spectrum while the antiresonances are characterised by

valleys in the magnitude spectrum.

2.3.2 *Properties of the Fourier transform Phase Spectrum* (FTPS) :

1. For any real x(n) the FTPS is an odd function of w.

2. For any x(n) the phase spectrum is the wrapped phase function, i.e. values of the phase function are restricted to $\pm\pi$.

3. If the signal x(n) is shifted by n_0 in the time domain, a linear phase component $e^{-j\omega n_0}$ is added to the FTPS for all ω.

4. The IFT of $e^{j\theta(\omega)}$ gives an ***all-pass signal.***

5. If a signal x(n) is the impulse response of a cascade of resonators and antiresonators, the overall FTPS of x(n) is obtained as : the wrapped phase spectrum of (the sum of the unwrapped phase spectra of individual antiresonators - the sum of the unwrapped phase spectra of resonators). Owing to the additive property, the resolution available in the phase spectrum is generally higher than the resolution available in the magnitude spectrum.

2.4 ***Relationship between the Spectral Magnitude and Phase of a Signal through Group Delay functions:***

Although the FT magnitude and phase spectra are independent functions of frequency, there are certain conditions under which the two are related. In some situations in communication engineering (as in sensor array imaging for example) it is possible that (a) either the magnitude or phase spectrum is available, (b) one of the spectra is corrupted by noise and is hence unreliable or (c) a few of the samples are missing due to some faults in the receiving elements (B.Yegnanarayana, C.P.Mariadassou and Pramod Saini; 1990, B.Yegnanaryana. S.T.Fathima and Hema A. Murthy,19871.

In such situations it may be useful to know the relationship between the FT magnitude and phase spectra, inorder that one of the spectra can be obtained from the other. Once both the spectra are

available, the time domain signal can be estimated.

Before we derive the relationship between the FT magnitude and the FT phase spectrum we define a classification of signals based on the roots of the z-transform of the signal. These definition are required to bring out the relationship between the Magnitude and Phase spectra.

The z-transform of a finite length signal $x(n)$ can be represented by the mapping (Eqs (2.1):

$$\mathcal{Z} : S_1 \rightarrow S_2 \Rightarrow X(z) = \sum_{n=0}^{N-1} x(n)z^{-n} = \frac{N(z)}{D(z)} \qquad (2.9)$$

where

$N(z) = (z - z_{m0})(z - z_{m1}) \ldots (z - z_{mr})$, r is the degree of the polynomial $N(z)$

and

$D(z) = (z - z_{d0})(z - z_{d1}) \ldots (z - z_{dp})$, p is the degree of the polynomial $D(z)$

then S_{min} the set of all minimum phase signals is defined by

$$S_{min} = \{x : x(n) = \mathcal{Z}^{-1}(X(z)) : \forall i \ |z_{mi}| < 1, \ |z_{di}| < 1\}$$

and S_{max} the set of all maximum phase signals is defined by

$$S_{max} = \{x : x(n) = \mathcal{Z}^{-1}(X(z)) : \forall i \ |z_{mi}| > 1, \ |z_{di}| > 1\}$$

and S_{mix} the set of all mixed phase signals is defined by

$$S_{mix} = \{x : x(n) \in S - (S_{min} \cup S_{max})\}$$

where S is the set of all signals.

For a discrete time signal $\{s(n)\}$ the DTFT is defined as

$$S(\omega) = \sum_{n=0}^{m} s(n)e^{-jwn}$$

$$= |S(\omega)|ej\theta(\omega) \qquad (2.10)$$

where $\theta(\omega)$ is the unwrapped phase function and $|S(\omega)|$ is the magnitude function. If the z-transform $S(z)$ of $s(n)$ does not have

any zeros on the unit circle, the continuity of $|S(\omega)|$ is guaranteed on the unit circle and we can define the complex cepstrum [A.V.Oppenheim and R.W.Schafer, Ch.10, 1975]

$$\hat{s}(n) \quad \overset{\mathcal{F}}{} \quad \hat{S}(\omega) = \ln|S(\omega)| + j\theta(\omega).$$

If $\hat{s}(n)$ is causal then the real and imaginary parts of $\hat{S}(\omega)$ corresponding $\ln|S(\omega)|$ and $\theta(\omega)$ are related through the Hilbert transform [A.V.Oppenheim and R.W.Schafer, Ch.7, 19751

$$\theta(\omega) = -\frac{1}{2\pi} \oint_{-\pi}^{\pi} \ln|S(\omega)| \cot\left[\frac{\omega - \Omega}{2}\right] d\Omega \qquad (2.11a)$$

$$\ln|S(\omega)| = \hat{s}(0) - \frac{1}{2\pi} \oint_{-\pi}^{\pi} \theta(\omega) \cot\left[\frac{\omega - \Omega}{2}\right] d\Omega \qquad (2.11b)$$

and

$$\hat{s}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln|S(\omega)| d\omega \qquad (2.11c)$$

This is nothing but the minimum phase condition [A.V.Oppenheim and R.W.Schafer, Ch.7, 19751 i.e. the zeros and poles of the z-transform S(z) lie within the unit circle. Alternative to Eqs(2.11) which relate the magnitude and the phase spectra of the Fourier transform, group delay functions can be used to relate the magnitude and phase spectra. This is the topic of the next Section in this Chapter.

2.4.1 :Group Delay functions :

Definition : If the phase spectrum ($\theta(\omega)$) of a signal is defined'as a continuous function of ω, the group delay function is defined as

$$\tau(\omega) = -\frac{d(\theta(\omega))}{d\omega} \qquad (2.12)$$

The deviation of the group delay function away from a constant indicates the degree of nonlinearity of the phase. The group delay function is expressed in seconds.

Let the Fourier transform $V(\omega)$ of a minimum phase signal $\{v(n)\}$

be represented by

$$V(\omega) = |V(\omega)|e^{j\theta_v(\omega)} \qquad (2.13)$$

Then it can be shown that [A.V.Oppenheim and R.W.Schafer, Ch.10, 1975]

$$\ln|V(\omega)| = c(0)/2 + \sum_{n=1}^{\infty} c(n)\cos n\omega \qquad (2.14a)$$

and the unwrapped phase function

$$\theta(\omega) = \theta_v(\omega) + 2\pi\lambda(\omega) = -\sum_{n=1}^{\infty} c(n)\sin n\omega \qquad (2.14b)$$

where c(n) are the cepstral coefficients. A detailed description of the cepstrum and its properties can be found in [D.G.Childers, D.P.Skinner and R.C.Kemeriat; 19771.

Taking the derivative of Eq(2.14b) with respect to ω, we get

$$\theta'(\omega) = -\sum_{n=1}^{\infty} nc(n)\cos n\omega \qquad (2.15)$$

From the above equations (2.14) we note that for a minimum phase signal, the spectral magnitude and phase are related through the cepstral coefficients. Further the group delay function $\tau(\omega)$ can be obtained as the FT of the weighted cepstrum.

The group delay function can also be obtained directly from the discrete time signal as [A. V.Oppenheim and R.W.Schafer; 1975, Ch.7]

$$\tau(\omega) = \mathrm{Re}\left[\frac{FT(v(n))}{FT(nv(n))}\right] \qquad (2.16)$$

where Re stands for the real part. Therefore for minimum phase signals using the relations (2.14) and Eq.(2.16), the minimum phase signal can be obtained from its group delay function.

For mixed phase signals we require two sets of cepstral coefficients $\{c_1(n)\}$ and $\{c_2(n)\}$ for magnitude and phase functions separately as follows :

$$\ln|V(\omega)| = c_1(0)/2 + \sum_{n=1}^{\infty} c_1(n)\cos n\omega \qquad (2.17a)$$

and

$$\theta(\omega) = \theta_v(\omega) + 2\pi\lambda(\omega) = -\sum_{n=1}^{\infty} c_2(n)\sin n\omega \qquad (2.17b)$$

where $\{c_1(n)\}$ and $\{c_2(n)\}$ are the cepstral coefficients of the unique minimum phase signals derived from the spectral magnitude and phase respectively [B. Yegnanarayana, D.K Saikia and T.R. Krishnan, 19841.

Using Eqs. (2.17) two different group delay functions are defined

$$\tau_p(\omega) = \sum_{n=1}^{\infty} n c_1(n)\cos n\omega \qquad (2.18a)$$

and

$$\tau_m(\omega) = \sum_{n=1}^{\infty} n c_2(n)\sin n\omega \qquad (2.18b)$$

as the group delay function derived from the magnitude and phase respectively.

2.4.2 *Properties of Group delay functions*

(1) Poles (Zeros) of the transfer function show up as peaks (valleys) in the group delay domain (Appendix A).

(2) Additive property : Convolution of signals in the time domain is reflected as a summation in the group delay domain (Fig.2.3 and Appendix A).

(3) High resolution property : The resonance peaks (due to complex conjugate pairs of poles or zeros) of a signal are better resolved in the group delay domain than in the spectral magnitude (Fig.2.3). Furthermore the resonance information is confined to the narrow region around the pole or zero location as shown in Fig.2.3.

(4) For minimum phase signals

$$\tau_p(\omega) = \tau_m(\omega)$$

(5) For maximum phase signals

$$\tau_p(\omega) = -\tau_m(\omega)$$

(6) For mixed phase signals

$$\tau_p(\omega) \neq \tau_m(\omega)$$

(7) If a root is on the unit circle in the z-domain (say ω_o)

$$\tau_p(\omega_o) = \infty$$

(8) The group delay function does not suffer from the wrapping problen as it can be computed directly from the time.domain signal using Eq(2.18).

2.5 Digital Representation of Speech Signals

The notion of a representation of a speech signal is central to almost every area of speech communication. In this Section we briefly review the speech production process. We also discuss a model for speech production which is assumed in most speech

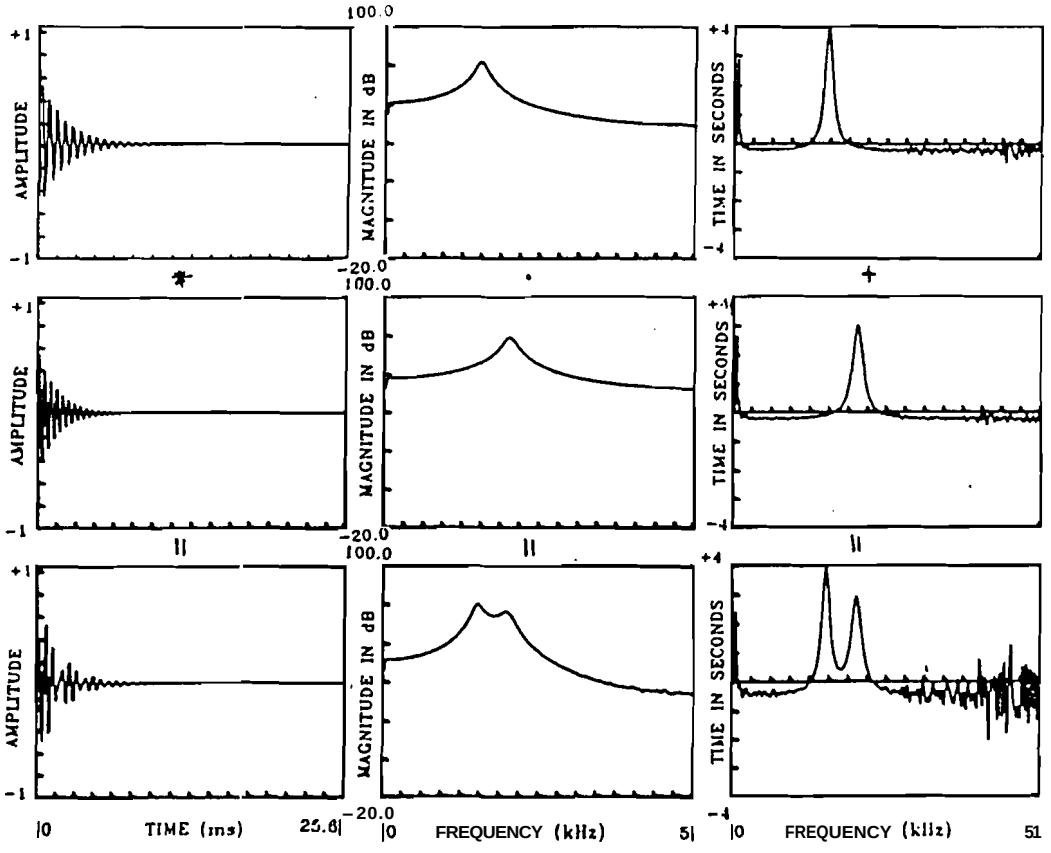


Fig.2.3 Illustration of the additive and high resolution . property of group delay functions. .

processing techniques. In some methods of speech processing this model is explicitly used to develop the methods of processing. In others although this model is not fundamental to the processing methodology developed yet this model is implicit in the methods developed for processing speech-like signals.

2.5.1 *A digital Model for Speech Production*

A schematic diagram of the human vocal apparatus is shown in Fig.2.4 [reproduced from W.A.Ainsworth; Ch.2, 1988]. The vocal tract is an acoustic tube that is terminated at one end by the vocal cords and at the other end by the lips. An ancilliary tube, the nasal tract, can be connected or disconnected by the movement of the velum. The shape of the vocal tract is determined by the position of the lips, jaw, tongue and velum.

Sound is generated in this system in three ways. Voiced sounds are produced by exciting the vocal tract with quasiperiodic pulses of air pressure caused by vibration of the vocal cords. Fricative sounds are produced by forming a constriction somewhere in the vocal tract, and forcing air through the constriction, thereby creating a turbulence which produces a source of noise to excite the vocal tract. Plosive sounds are created by completely closing off the vocal tract, building up pressure and then quickly releasing it. All these sources create a wide band excitation of the vocal tract which in turn acts as a linear time varying filter which imposes its transmission properties on the frequency spectrum of the sources. The vocal tract can be characterised by its natural frequencies (or formants) which correspond to resonances in the sound transmission characteristics of the vocal tract.

A typical speech waveform is shown in Fig.2.5 which illustrates some of the basic properties of the speech signal. We see for

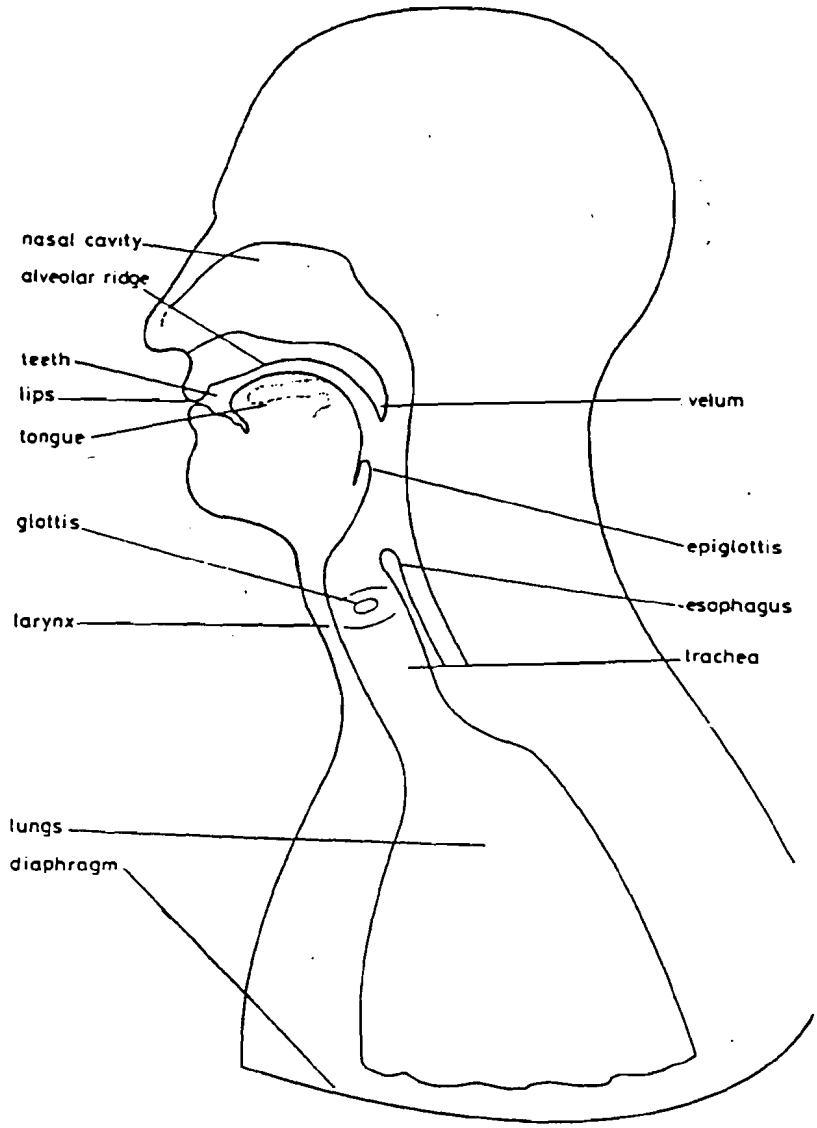


Fig.2.4 Articulators used in the production of speech sounds.

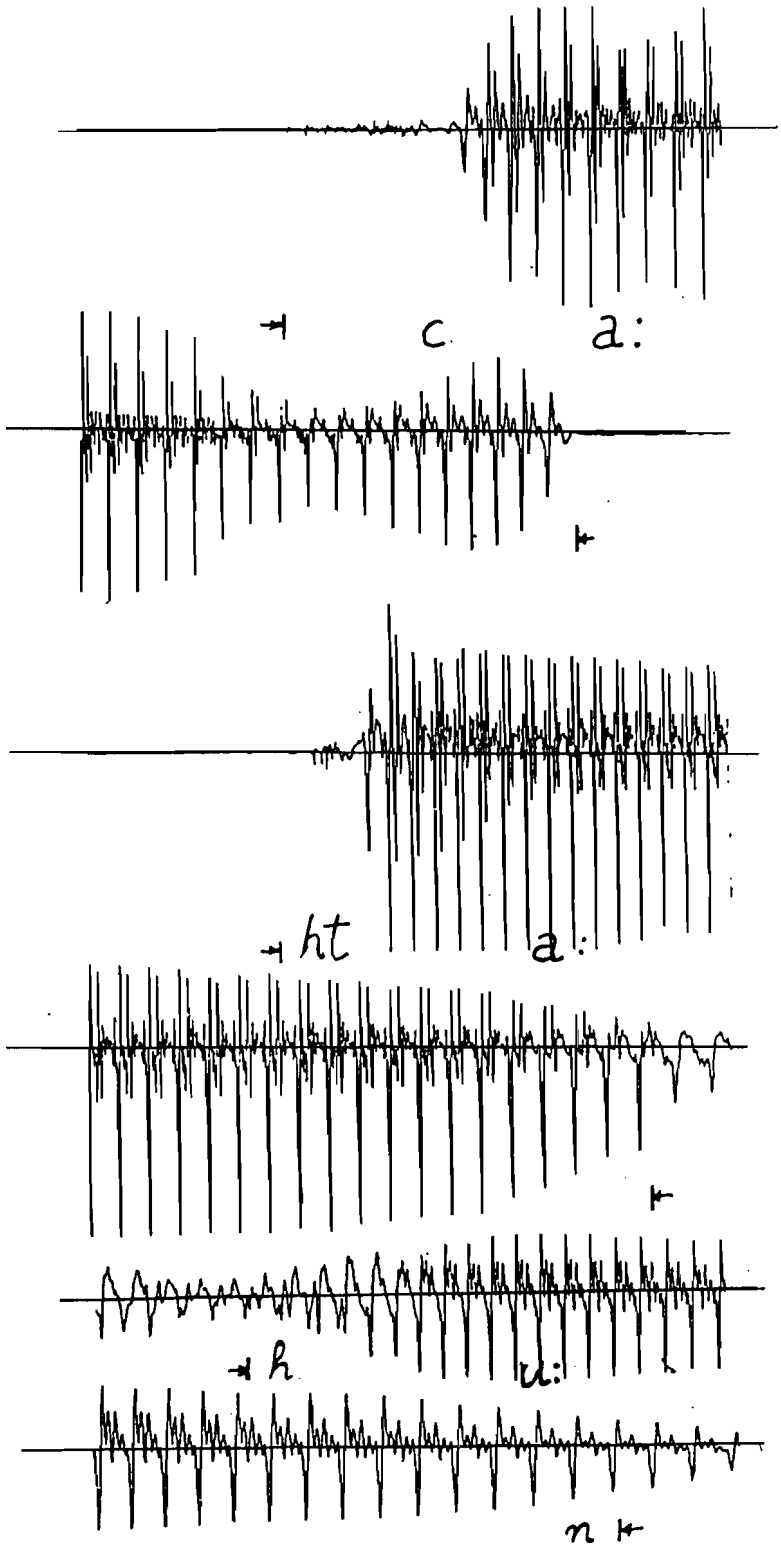


Fig.2.5 An illustration of a speech waveform corresponding to the utterance "ca: hta: *hu:n*".

example, that although the properties of the waveform change with time, it is reasonable to view the speech waveform as being composed of segments during which the signal properties remain rather constant. Such segments are demarked in Fig.2.5 below the waveform. These sample segments have the appearance of a low level random (unvoiced) signal (as in c or t in Fig.2.5) or a high level quasi periodic (voiced signal) (as in a: or u:) with each period displaying the exponential decaying response properties of an acoustic transmission system. We note that the dynamic range of the waveform is large, i.e., the peak amplitude of a voiced segment is much larger than the peak amplitude of an unvoiced segment.

Because the sound source and vocal tract shape are relatively independent, a reasonable approximation is to model them separately as shown in Fig.2.6 [R.W.Schafer and L.R.Rabiner; 1978]. In this digital model, samples of the speech waveform are assumed to be the output of a time-varying digital filter that approximates the transmission properties of the vocal tract and the spectral properties of the glottal pulse shape. Since, as is clear in Fig.2.5 the vocal tract shape changes rather slowly in continuous speech (likewise its sound transmission properties) it is reasonable to assume that the digital filter in Fig.2.6 has fixed characteristics

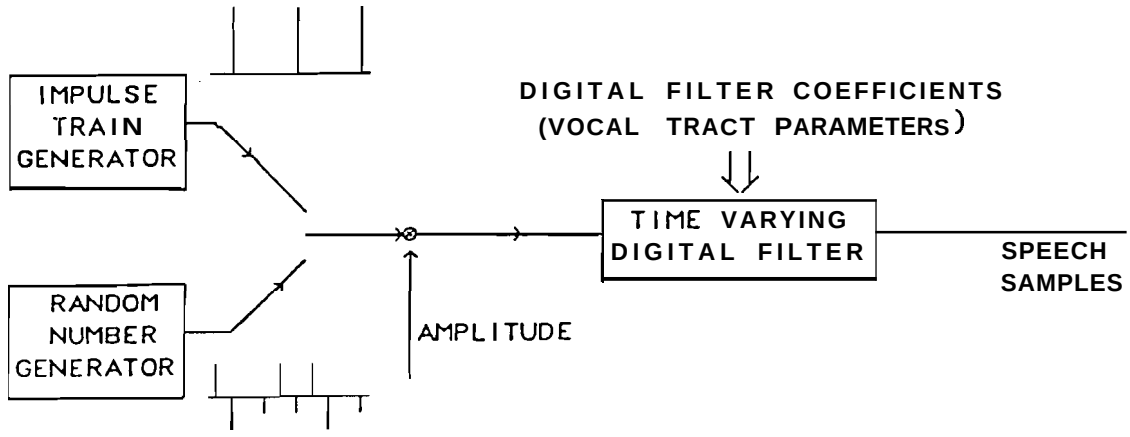


Fig.2.6 A digital model for speech production.

over a time interval of the order of 10ms. Thus the digital filter may be characterised in each interval by an impulse response or a set of coefficients for a digital filter. For voiced speech the digital filter is excited by an impulse train generator that creates a quasiperiodic impulse train in which the spacing between impulses corresponds to the fundamental period of the glottal excitation. For unvoiced speech the filter is excited by a random number generator that produces a flat spectrum of noise. In both cases an amplitude control regulates the intensity of the input to the digital filter.

This model is the basis for a wide variety of representations of speech signals. These are conveniently characterised as waveform representations or parametric representations depending upon whether the speech waveform is represented directly or whether the representation is in terms of time-varying parameters of the basic speech model. In the forthcoming discussion on speech processing we assume this simplified model for speech production.

2.6 *Speech processing*

In the model for speech production discussed in the previous Section it is clear that parameters that are required to be extracted from the speech signal are the system and source parameters for applications in (a) synthesis of speech and (b) recognition of speech. The vocal tract changes shape for the articulation of new sounds. These changes in shape are characterised by the change in the parameters of the digital filter in Fig.2.6. The closing and opening of the glottis and the vibration of the vocal cords are described by parameters that characterise the source. The vocal tract is also described in terms of its resonances (or formants) which may be derived either from the model parameters or from the

speech signal spectrum. The source information may be derived by passing the speech signal through the inverse of the model system. In general, the model system is derived so as to represent the smoothed short-time spectrum of speech. The fine structure of the spectrum is used to derive the excitation information. All-pole or pole-zero models are usually assumed for the vocal tract system. Linear prediction analysis is an effective method for determining the parameters of an all pole model for the speech signal [J.Makhoul, 1975]. Cepstrum analysis [A.V.Oppenheim and R.W.Schafer; 19631 is another effective method for determing the formant and pitch information from the speech signal. In this section we briefly discuss some of the existing methods for the analysis of speech signals to obtain parameters of the model that characterise the system and source information.

2.6.1 *Estimation of System Parameters from Speech* :

Experiments in the analysis and perception of speech [J.L.Flanagan; 1956. J.L.Flanagan and L.Cherry; 19691 have shown that certain speech sounds, notably the vowels may be identified and synthesised primarily from a knowledge of the formant frequencies. The formant frequencies, therefore, appear to be important information bearing elements of speech. In fact, analytical analysis of the vocal mechanism [G.Fant; 19591 has shown that the acoustic output during vowel production may be specified rather accurately from a knowledge of formant frequencies and the fundamental vocal frequency.

Three effects are of major importance in limiting the accuracy with which formant frequencies and bandwidths of vowels may be estimated from spectral data [E.N.Pinson; 19631. These are (1) the effect of source periodicity on the spectrum, (2) the effect of the

source spectrum envelope and (3) the effect of time averaging over both closed glottis and open glottis condition.

Because of the periodicity of the source (i.e. the puffs of air flowing through the glottis that excite the vocal tract), the spectrum of the acoustic waveform consists roughly of lines at a fundamental (pitch) frequency and its harmonics (Fig.2.7). Little

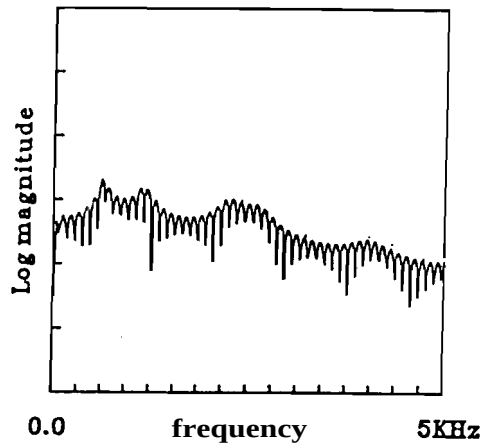


Fig.2.7 Illustration of the manifestation of periodicity in the time waveform as fine structure on the spectrum.

information is available about the spectrum between the pitch harmonics so the frequencies of the spectral peaks must be interpolated between these lines. A further complication arises if the envelope of the source spectrum exhibits rapid variations with frequency. Since the measured spectrum is the product of the spectrum of the glottal source and the spectrum of the transfer function of the vocal tract, the effect of the source spectrum is to distort the features of interest. Finally, the effect of relatively long averaging times used in any spectrum measurement technique (Fourier analysis, for instance) reduces the accuracy with which the formants may be obtained. In this context the output at any time depends upon both open and closed glottis portions of the speech waveform, whereas the vocal tract resonance characteristics are different in both these intervals.

Several methods have been used to estimate the formant frequencies for voiced speech data. Almost all these techniques have as a common starting point, the transformation of the acoustic data into spectral form. The methods used to obtain the spectrum have included use of sound spectrograph , bank of bandpass filters and pitch synchronous Fourier analysis.

Most of the techniques that exist for estimating the vocal tract parameters from the speech signal can be classified as model-based and non model-based techniques. The technique by Flanagan [J.L.Flanagan; 19561 describes an analog approach to estimating formants in continuous speech. This paper describes the use of peak-picking on a short-time Fourier representation to obtain an estimate of Formant frequencies. The paper by Dunn [H.K.Dunn; 1961] deals with the use of a sound spectrograph for vowel bandwidth measurements. Schafer and Rabiner [R.W.Schafer and L.R.Rabiner; 19701 describe homomorphic processing to the estimation of the vocal tract frequency response. Quatieri [T.F.Quatieri, Jr., 19791 has developed a technique for the improvement of speech analysis/synthesis systems using homomorphic deconvolution in which both a minimum and maximum phase reconstruction is addressed. Another digression from the traditional homomorphic deconvolver is the work due to Verhelst and Steenhaut [W.Verhelst and O.Steenhaut; 19861. In this approach a complex model for homomorphic deconvolution is suggested in which an approximation to the influence of window length is included. Also the spectral sampling inherent in voiced speech is explicitly represented. These techniques come under the category of nonmodel-based techniques

Most of the recent methods for the estimation of the vocal tract parameters are based on a model-based approach.

The work by [B.S.Atal and S.L.Hanauer; 19711, [F.Itakura and S.Saito; 19701, [J.D.Markel,1972a], McCandless [S.S.McCandless; 19741, [B.S.Atal and M.R.Schroeder; 1979], [E.Denoel and J.-P.Solvay; 19851, [A.El-Jaroudi and J.Makhoul; 1987], [C.H.Lee; 1987,19881 and [C.Duncan and M.A.Jack, 1988] are concerned with linear predictive analysis methods for estimating the vocal tract function and formant frequencies.

In the work by Atal and Hanauer [B.S.Atal and S.L.Hanauer; 19711 linear prediction analysis [J.Makhoul; 19751 is used to estimate the time-varying parameters of the speech wave, namely the prediction coefficients and pitch. Itakura and Saito [F.Itakura and S.Saito; 19701 discuss a maximum likelihood approach to the estimation of the linear prediction coefficients. Markel [J.D.Markel; 1972a] discusses an algorithm based on the digital inverse filter formulation for formant trajectory estimation. McCandless [S.S.McCandless; 19741 suggests a method for formant extraction from linear prediction spectra to take into account spurious peaks, merged peaks, etc. Atal and Schroeder [B.S.Atal and M.R.Schroeder; 19791 use a subjective error criterion to estimate the LP coefficients. Denoel and Solvay [E.Denoel and J.-P.Solvay; 1985] modify the error criterion in standard linear prediction to estimate the LP coefficients that characterise the vocal tract. An absolute error criterion rather than the usual squared error criterion is used.

Song etal. [K.H.Song and C.K.Un; 19831 discuss a method for Pole-zero modeling of speech using a higher order pole model fitting and decomposition method. The work by [H.Morikawa and H.Fujisaki; 19841 is based on a state space representation for the speech production process. In this technique speech is modeled as an ARMA (auto-regressive moving average) process with variable order.

Kopec [G.E.Kopec; 1986a, 1986b] uses a hidden Markov model and vector quantisation to track formants accurately. Other approaches include maximum likelihood spectral estimation and its application to speech analysis [M.J.McAulay; 19841. Yet another work due to McAulay etal. [M.J.McAulay and T.F.Quatieri; 19861 is based on a sinusoidal representation of speech. Regoll [G.Regoll; 19861 describes a new algorithm based on an extended Kalman filter model for the time-varying digital filter in the model for speech production. Schroëter etal. [J.Schroëter, J.N.Larar and M.M.Sondhi; 19871 use a vocal tract/cord model for parameter estimation from speech. Lee [C.H.Lee; 1987, 19891 develops an algorithm for linear prediction, in which the sum of appropriately weighted residuals is minimised to estimate the LP coefficients. In the work by [G.Duncan and M.A.Jack; 19881 a pole-focussing approach is taken to estimate the LP coefficients in the filter that characterises the vocal tract

Although the methods described in these papers take different points of view to formulating the analysis methods, the resulting methods have much in common and results obtained are comparable in that the time complexity of the algorithm is proportional to the resolution that can be obtained.

We now briefly describe two commonly used methods for formant estimation, one based on homomorphic processing (a nonmodel-based technique, also called cepstrum analysis) and the other based on linear prediction analysis (a model-based technique).

2.6.1.1 Cepstrum Analysis

The technique of cepstral processing is used for separating the excitation signal of a speech wave from the filter part. This makes it easier to estimate both the periodicity of the excitation and the frequencies of the formants.

Based on the model of Fig.2.6 the speech waveform s(n) is obtained as the convolution of the excitation signal e(n) and the impulse response of the digital filter h(n).

$$\mathcal{F}[s(n)] = \mathcal{F}[e(n)]\mathcal{F}[h(n)] \qquad (2.19a)$$

$$S(\omega) = E(\omega)H(\omega) \qquad (2.19b)$$

The logarithm of the transform is obtained as

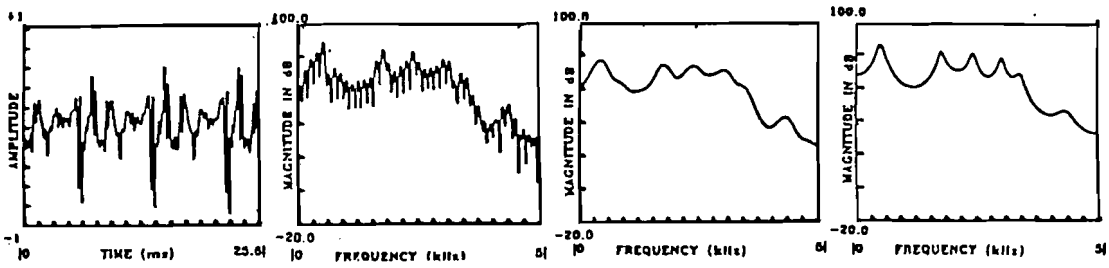
$$\log S(\omega) = \log E(\omega) + \log H(\omega) \qquad (2.20)$$

Finally the inverse DFT of the transform is computed

$$\mathcal{F}^{-1}[\log S(\omega)] = \mathcal{F}^{-1}[\log E(\omega)] + \mathcal{F}^{-1}[\log H(\omega)] \qquad (2.21)$$

The resulting spectrum of the log of the frequency spectrum is called the ***cepstrum.*** The horizontal axis of a cepstrum has the dimensions of time and is called ***quefrency.***

The result of this complex transformation is to .increase the effects of the fundamental relative to other frequencies present in the original waveform. Consequently, the cepstrum will contain a large peak corresponding to the fundamental. The position of this peak can be used to obtain an estimate of the fundamental frequency. If this peak is removed by cutting of the cepstrum from below this quefrency, the ripple on a spectrum caused by the effects of the fundamental can be reduced. Taking a DFT of cepstrum with the fundamental removed results in a smoothed spectrum. The peaks of this spectrum are identified as formants (Fig.2.8c).



(a) signal (b) FT magnitude *spectrum* (c) Cepstrally *smoothed spectrum* (d) LP *spectrum*

Fig.2.8 Ceptrum and Linear Prediction methods of estimating the, smoothed spectra for a segment of speech.

2.6.1.2 Linear Prediction analysis

The foregoing analysis of the speech signal based on the cepstrum makes no assumption about how the speech was produced. Linear prediction analysis, however assumes that the signal being analysed is produced by passing an excitation signal through a suitable filter. This is a good model for the production of many speech sounds. Hence this is an appropriate technique for speech signal analysis.

Suppose that a waveform x(t) has been digitised. If x(t) was continuous, then the current sample x(n) can be predicted from the previous sample :

$$x(n) = a_1 x(n-1) + e(n) \qquad (2.22)$$

where the coefficient a_1 is obtained so as to minimise the error signal e(n). This idea can be extended by predicting x(n) from the last p samples:

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + e(n) \qquad (2.23)$$

The coefficients a_k are obtained by minimising the error $E[e(n)^2]$ where E stands for the expectation operator. This leads to a set of simultaneous equations in the a_ks :

$$\sum_{k=1}^{p} a_k \, E[x(n-j)x(n-k)] = E[x(n)x(n-j)], \quad j = 1,2,\ldots,p \qquad (2.24)$$

Replacing E[ ] by time averages and defining

$$R(m) = \sum_{n} x(n)x(n+m) \qquad (2.25)$$

leads to the autcorrelation method of LP analysis. In practice, the signal is not known over an infinite range, so as usual it is windowed. Replacing E[ ] by the time average $\sum_{n=h}^{h+N-1}$ leads to the covariance method of LP analysis

$$\sum_{k=1}^{p} a_k \sum_{n=h}^{h+N-1} x(n-j)x(n-k) = \sum_{n=h}^{h+N-1} x(n)x(n-j), \quad j=1,2,\ldots,p \qquad (2.26)$$

Once the a_ks are obtained the smoothed spectrum can be obtained by computing

$$S(\omega) = \left| \frac{1}{1 - \sum_{k=1}^{p} a_k e^{-j\omega k}} \right|^2 \qquad (2.27)$$

The peaks of this spectrum correspond to formants as shown in Fig.2.8d.

2.6.2 *Estimation of Source parameters*

Accurate and reliable measurement of the pitch period of a speech signal from the speech waveform alone is often exceeding difficult for several reasons. One reason is that the glottal excitation waveform is not a perfect train of periodic impulses. Although finding the period of a perfectly periodic waveform is straightforward, measuring the periodicity of a speech waveform. which varies both in period and in the detailed structure of the waveform within a period can be quite difficult. A second difficulty in measuring pitch period is the interaction between the vocal tract and the glottal excitation. In some cases the formants of the vocal tract can alter the structure of the glottal waveform significantly. This occurs when the articulatory movements are very rapid. This causes rapid changes in formants. A third problem in reliably measuring pitch is the inherent difficulty in defining the exact beginning and end of each pitch period during voiced speech segments. For example, based on the acoustic waveform alone, some candidates for defining the beginning and end of the period include the maximum value during the period and zero crossing prior to the maximum. The only requirement on such a measurement is that it be consistent from

period to period. Fig.2.9 shows possible estimates of pitch period. Both measurements are likely to give different values for the pitch period. The pitch period discrepancies are due not only to the quasiperiodicity of the speech waveform, but also the fact that peak measurements are sensitive to the formants, noise and any dc level in the waveform.. A fourth difficulty in pitch detection is distinguishing between unvoiced speech and low level 'voiced speech. In many cases transitions between unvoiced speech and low level voiced speech are very subtle and are thus extremely hard to pin point.

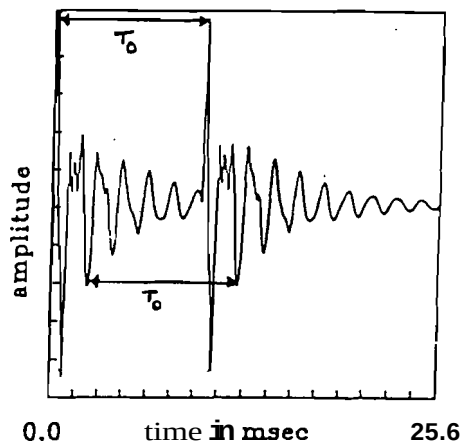


Fig.2.9 Possible estimates of pitch period.

Types of pitch detectors

As a result of the numerous difficulties in pitch measurements. a wide variety of sophisticated pitch detection methods have been developed. Basically a pitch detector is a device which makes a voiced-unvoiced decision and, during voiced speech, provides a measurement of the pitch period. However, some pitch detection algorithms just determine the pitch during voiced segments of speech and rely on some other technique for the unvoiced-voiced

decision. Pitch detection algorithms can be roughly divided into the following categories:

1) A group which utilises principally the time domain properties of speech signals.

2) A group which utilises principally the frequency domain properties of speech signals.

3) A group which utilises both the time and frequency domain properties of speech signals.

Time domain pitch detectors operate directly on the speech waveform to estimate the pitch period. For these pitch detectors the measurements most often are peak and valley measurements, zero crossing measurements, and autocorrelation measurements. The basic assumption that is made in all these cases is that if a quasiperiodic signal has been suitably processed to minimise the effects of the formant structure, the simple time domain measurements will provide good estimates of the period. Some of the pitch detection algorithms that belong to this category are (a) Modified autocorrelation method [J.J.Dubnowski, R.W.Schafer abd L.R.Rabiner; 1976], (b) Average magnitude difference function [M.J.Ross, H.L.Shaffer, A.Cohen, R.Freudberg and H.J.Manley; 19741 (c) Data Reduction Method [N.J.Miller; 19751 and (d) Parallel processing method [B.Gold and L.R.Rabiner; 19691.

The modified autocorrelation pitch detector is based on the center clipping method due to Sondhi [M.M.Sondhi; 19681. In this method the signal is low pass filtered, the low pass filtered signal is then sampled at 10kHz, sectioned into overlapping 30ms sections for processing. A clipping level is chosen, the section is center clipped and peaks of the autocorrelation function of this signal correspond to pitch period.

The class of frequency domain pitch detectors use the property that if the signal is periodic in the time domain, then the frequency spectrum of the signal will consist of a series of impulses at the fundamental frequency and its harmonics. Thus simple measurements can be made on the frequency spectrum of the signal (or a nonlinearly transformed version of it as in the cepstral detector,) [A.M.Noll; 1967, R.W.Schafer and L.R.Rabiner; 19701. In the cepstral detector due to Noll, the cepstrum is computed. The cepstrurn has a strong peak corresponding to the pitch period. The peak is located in the cepstrum and taken as an estimate of the pitch period.

The class of hybrid pitch detectors incorporates features of both the time and frequency domain approaches to pitch detection. For example, a hybrid pitch detector might use frequency domain techniques to provide a spectrally flattened time waveform, and then use autocorrelation measurements to estimate the pitch period as in SIFT [J.D.Markel; 1972b] and Spectral Equalisation LPC method using Newton's transformation [B.S.Atal and L.R.Rabiner; 19761.

In the SIFT technique, the given signal is low pass filtered, the inverse filter coefficients for this signal are computed. The inverse filter output is then obtained. The autocorrelation of this sequence is computed. The largest peak within specified limits corresponds to the pitch period. A comparative of study of some of the standard pitch detection algorithms can be found in [L.R.Rabiner, M.J.Cheng, A.E.Rosenberg and C.A.McGonegal; 1976].

Recently Gong and Haton [Y.Gong and J.P.Haton; 1987] have brought about a new formulation for the pitch estimation problem. The speech signal is modeled as a sequence of a specified function in a time-dependent manner which allows the period and amplitude of the excitation signal to be time-varying. Andrews et al. [M.S.Andrews,

J.Picone and R.D.Degroat; 1990] introduce a cepstrum based pitch estimator which couples the signal enhancement capabilities of MUSIC [S.M.Kay; 1988] with the harmonic spectrum estimation capabilities of cepstrum. Hodgson et al. [L.Hodgson, M.E.Jernigan and B.L.Wills; 1990] develop a new algorithm for cepstrum pitch detection where a nonlinear model for the vocal tract is assumed. Slaney's [M.Slaney; 1990] perceptual pitch detector combines a cochlear model with a bank of autocorrelators. An independent autocorrelation is performed for each channel, the information is combined to obtain an estimate of pitch period.

2.6.3 The problem of Speech Enhancement

In most practical situations signals are contaminated by noise. Different approaches may be needed to deal with different types of noises such as quantisation, multiplicative, convolutional, signal-dependent and additive. The topic that is addressed in this thesis is the problem of processing noisy speech where the noise that is considered is additive. Even in this limited context, there are a variety of situations in which speech enhancement is desired [J.S.Lim; 1979b]. In speech processing we observe that accurate information about (a) the vocal tract resonances and (b) the excitation is essential for synthesising speech of high quality.

The classical work of Wiener and others gives an approach for deriving an optimal filter that tends to suppress the noise while retaining the desired signal unchanged [S.S.Haykin; 1986, B.R.Widrow and Stearns; 1986]. The basic assumption in Wiener filter theory is that both signal and noise are stationary, which is seldom true in the context of speech. Nevertheless several techniques based on the approximation of the optimum Wiener filter using tapped delay line have been suggested in the literature. An approach based on the

concept of the adaptive noise canceller due to Sambur [M.R.Sambur;1978] uses the Least Mean Squares adaptive filtering approach to remove the effects of additive noise on the speech signal. The logic that is used in this approach is that voiced portions of the speech signal are periodic and a frame of this portion delayed by a few pitch periods will be highly correlated while the noise will be uncorrelated. Another approaah based on the is adaptive comb filtering [J.S.Lim, A.V.Oppenheim and L.D.Braida; 1978b]. The property that is exploited here is that the energy of a periodic waveform is concentrated in bands of frequencies. Unfortunately, techniques based on this approach seldom succeed bacause neither is the noise stationary nor is the sample representative of the noise in the system. Although Sambur has obtained satisfactory results we are still left with the basic problem of identifying the voiced and unvoiced portions of a speech signal and estimating the periodicity in the waveform.

A particular class of speech enhancement systems are based on the assumption that the short time magnitude spectrum of speech is more important than the short time phase spectrum. In such systems an estimate of the magnitude spectrum is first made and then combined. with that of the phase spectum to produce enhanced speech. The most commonly employed procedure is to estimate the noise power spectrum and then use spectral subtraction. This requires that some knowledge about the statistics of the noise is available[S.F.Boll; 1979, S.F.Boll and D.C.Pulsipher; 1980].

Other approaches to speech enhancement make use of the underlying model for speech production. Model based approaches for speech enhancement estimate the parameters of the model rather than the speech signal. This information is then used in a speech

analysis/synthesis system. Homomorphic deconvolution is one method for estimating the impulse response of the speech production model. Later systems attempt to model the vocal tract system as accurately as possible. The problem of estimating the parameters of the model for speech production has been dealt with extensively in the literature. Although the estimation of these parameters is straightforward for clean speech, it is rather difficult for noisy speech. One such approach (in fact the most successful) is due to Lim [J.S.Lim and A.V.Oppenheim, 1978a] where a maxmimum aposteriori estimation procedure is employed to estimate these parameters. In the context of noise a set of nonlinear equations are obtained in place of Eq (2.24) for autoregressive parameter estimation. Although the system suggested is suboptimal, nevertheless the speech output is good.

Another class of algorithms enhance speech in various contexts by changing the time scale of speech, i.e. slowing it down or speeding it up. Malah [D.Malah; 19791 presents a method in which the speech signal is decomposed into complex exponentials, then the frequency of each exponential is modified by the same ratio in each channel, without affecting the duration and amplitude of the exponential. The resulting speech is obtained by combining these exponentials has the same duration but all frequency components are scaled. Portnoff [Portnoff; 1981a, 1981b] presents a method in which the short time Fourier transform of speech is modified and speech synthesised from the modified Fourier transform. In this kind of approach a particular frequency that is important for intelligibility can be independently controlled.

2.7 The *problem of Spectrum Estimation*

Spectrum analysis of signals is performed to extract the

information about the system that generated the signal. Since the signal available for analysis is usually of short duration and also noisy, one can only attempt to estimate the spectrum or the system characteristics, rather than compute the spectrum, The accuracy of the estimated spectrum depends on the bias and variance of the estimate, which in turn depends on the nature of the signal, its duration type of windowing and noise.

The main issue in spectrum estimation is to obtain a high resolution from short data record and from data combined with noise. Effects of short data records, windowing, noise and model order have been studied extensively in [S.M.Kay; 1988, S.L.Marple; 19871. In all these cases two classes of problems are addressed (i) estimation of autoregressive parameters in noise and (ii) estimation of component sinusoids from finite duration noisy data.

Although model-based techniques [D.G.Childers; 1978, S.B.Kesler; 19861 are gaining popularity over that of nonmodel based techniques, it is well known that the biggest drawback of these techniques is that they fail miserably when no apriori information is available either about the signal of interest or noise. In the absence of any apriori information, the Fourier-based methods [C.Bigham, M.D.Godfrey and J.W.Tukey; 1967, P.D.Welch; 19671 are the most successful techniques for estimating parameters from the spectrum.

2.8 Motivation for the Current Research

It is interesting to notice that all the frequency domain techniques discussed so far (both in the context of speech analysis and spectrum estimation) whether model based or otherwise use the magnitude spectrum as the starting point. The phase spectrum of the signal is neither modelled nor estimated. In fact. the phase of the signal is not considered at all. This is perhaps due to the

difficulties encountered in processing the phase

In many situations it is observed that the phase spectrum of the signal rather than the magnitude spectrum of the signal is important for preserving the important features of the signal. This observation has been made in a number of situations, namely, acoustic holography, X-ray crystallography and image analysis [A.V.Oppenheim and J.S.Lim; 19811. Several algorithms have been developed to reconstruct a signal from either Fourier transform phase or magnitude. Among them are the algorithms for signal retrieval from phase developed by [M.H.Hayes, J.S.Lim and A.V.Oppenheim; 1980, P.L.Vanhove and M.H.Hayes; 1983, J.R.Fienup; 1987, Thomas R.Crimmins; 1987, N.Nakajima; 1987, S.L.Curtis and A.V.Oppenheim; 19871. In the context of speech, the quality and intelligibility of an utterance are completely restored when the phase of the signal and a flat magnitude spectrum are used as initial estimates in an iterative algorithm [B.Yegnanarayana, S.T.Fathima and Hema A. Murthy, 1987]. Although the phase contains all the information relating to events, namely edges in an image, formant transitions in speech and linear phase in both one-dimensional and two-dimensional signals, it is difficult to capture this information directly from the phase because it appears to be noisy and difficult to interpret.

Manipulation of the Fourier transform Phase directly for feature extraction requires that the phase spectrum of the signal be first of unwrapped. Some algorithms for unwrapping of Fourier transform phase [Tribolet; 1979, D.G.Ghiglia, G.A. Mastin and Louis A. Romero; 19871 are available for unwrapping the phase spectrum. These algorithms are computationally intensive and donot work for all classes of signals.

In this thesis, instead of directly processing the Fourier

transform phase spectrum of signals, we process the group delay functions of signals to estimate the features that characterise signals. The advantage of processing group delay functions rather than the phase function is that the group delay function has all the desirable properties of phase (additivity) and it does not suffer from the wrapping problem. The topic of the next Section is the group delay processing of signals.

2.9 *The Group Delay Approach to Signal Processing*

The group delay function does not suffer from the wrapping problem, but possesses all the desirable properties of the phase spectrum as was seen in the Section 2.4. Algorithms for computing the group delay functions as well as algorithms for deriving the signal from the group delay functions are given in [B.Yegnanarayana, D.K.Saikia and T.R.Krishnan; 19841. Reddy etal. [S.Reddy and M.N.S.Swamy; 19851 use the derivative of phase spectra to reduce the inherent windowing problem when using the DFT. [K.V.Madhu Murthy and B.Yegnanarayana; 19891 represents the first systematic study of the properties of group delay functions for the representation of signals. This study emphasises the usefulness of the representation of signals through group delay functions. Their observation is that the errors in representation can be reduced considerably by taking a large number of DFT points provided there are no roots on the unit circle in the z-transform of the signal. For any representation to be effective, it is desirable that the relevant information in the signal be preserved in that representation. If continuous frequency and time variables are used throughout there is no loss of information in any domain. But digital processing of data necessitates discretisation which may result in partial loss of information. We saw earlier that when a signal is represented by its

discrete version it was required that the signal be adequately sampled in order to avoid aliasing in the frequency domain. Similarly the discretisation of the signal may affect the accuracy of signal representation through group delay functions.

The properties of group delay functions can be exploited for many applications such as design of digital filters [B.Yegnanarayana; 1981a] and pole-zero modelling [B.Yegnanarayana; 1981b]. The properties of group delay functions allow manipulation of signal data effectively in many signal processing situations, like waveform estimation from an ensemble of noisy measurements [B.Yegnanarayana, J.Sreekanth and Anand Rangarajan; 19851. Most of the available literature on group delay functions make attempts to estimate signal data from the group delay function. Very few attempts have been made in which group delay functions or phase spectrum is used for parameter estimation [B.Yegnanarayana;1978, M.T.Manry; 19851 from signals. We now discuss some of the problems encountered when an attempt is made to estimate parameters from the group delay functions of signals. We first discuss the issues in group delay processing of speech signals. This is followed by a discussion of the application of group delay functions in spectrum estimation.

2.9.1 *Issues in Group Delay Processing of Speech signals*

For ease of explanation we assume the simple source filter model for speech production that was discussed in Section 2.5.

To illustrate the issues involved in group delay processing of speech we use the following filter model to represent the vocal tract system

$$V(z) = \prod_{k=1}^{s} \frac{1 - 2e^{-\pi B_k T}\cos(2\pi F_k T) + e^{-2\pi B_k T}}{1 - 2e^{-\pi B_k T}\cos(2\pi F_k T)z^{-1} + e^{-2\pi B_k T}z^{-2}} \quad (2.28)$$

This equation describes a cascade of digital resonators that have

unity gain at zero frequency. F_ks represent the frequencies of the resonators and B_ks represent the bandwidths of the resonators.

For voiced speech we assume that the filter is excited by a periodic train of impulses while for unvoiced speech, this filter is excited by random noise. Fig.2.10 shows the impulse response of the filter and its magnitude phase and group delay spectra, respectively. The information about the resonances appear as (i) peaks in the magnitude spectrum, (ii) phase transitions in the phase spectrum and (iii) as peaks in the group delay spectrum. Notice that the information about resonances in the group delay domain is concentrated around the peak.

Speech is produced by exciting the filter of Eq.(2.28) with a periodic impulse train or random noise. Fig.2.11 shows the impulse train and its group delay function. Fig.2.12 shows random noise sequence, and its group delay function. The combined response of Fig.2.10 and Fig.2.11 and its corresponding group delay are shown in Fig.2.13. This signal is an approximation to voiced speech. The combined response of Fig.2.10 and Fig.2.12 and its corresponding group delay is shown in Fig.2.14. This signal is an approximation to unvoiced speech. Notice that in the group delay domain, the information about the formants is completed masked by the group delay function corresponding to the source in both the figures (Fig.2.13 and Fig.2.14).

The zeros that are generated by the impulse train and finite window lie on the unit circle (Fig.2.15a) in the z-domain. The zeros that are generated by random noise are very close to the unit circle in the z-domain (Fig.2.15b). The poles due to the formants are well within the unit circle (Fig.2.15c). The group delay functions assume very large values at sampling points that are close to the zeros or

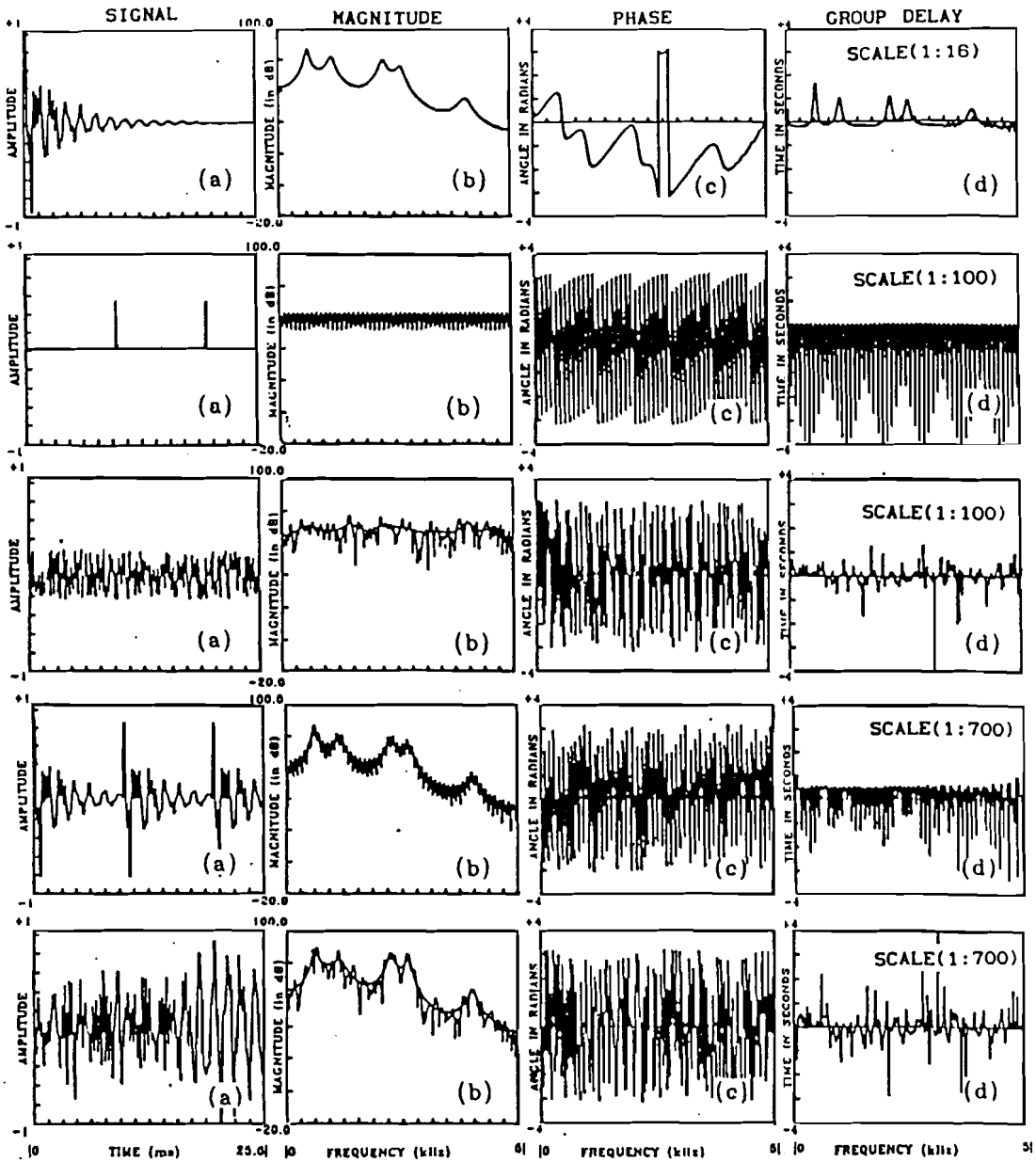


Fig.2.10 (a) Impulse response of an all-pole filter and its Spectra.

Fig.2.11 Impulse train and its Spectra.

Fig.2.12 Random noise and its Spectra.

Fig.2.13 Response of all-pole filter to impulse train and its spectra.

Fig.2.14 Response of all-pole filter to random noise and Spectra.

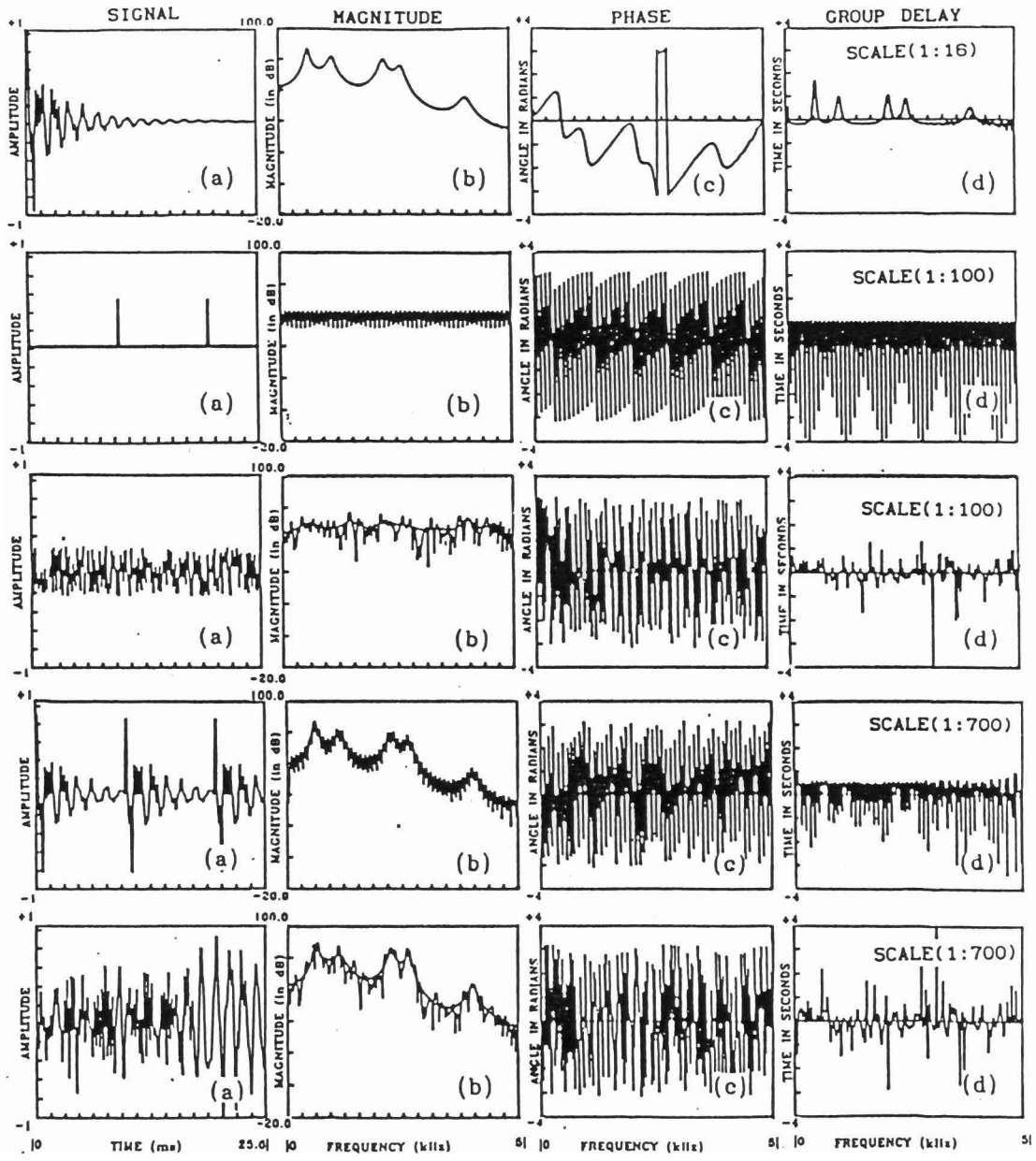
Fig. 2.10 (a) Impulse response of an all-pole filter and its Spectra.

Fig. 2.11 Impulse train and its Spectra.

Fig. 2.12 Random noise and its Spectra.

Fig. 2. 3 Response of all-pole filter to impulse train and its spectra.

Fig. 2.14 Response of all-pole filter to random noise and Spectra.



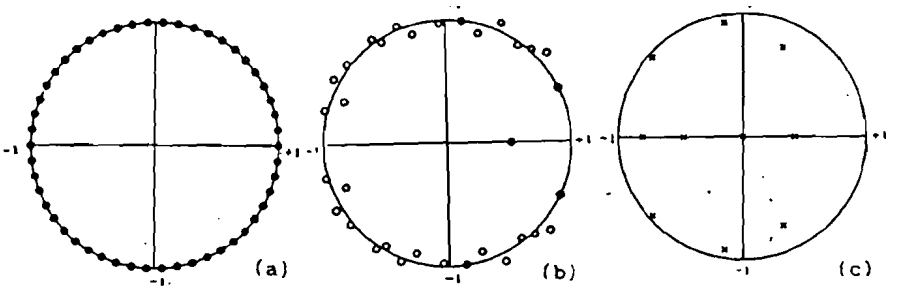


Fig.2.15 Distribution of roots in the z-plane for (a) impulse train (b) random noise and (c) all-pole filter.

poles. Since the group delay function is obtained by sampling the z-transform on the unit circle, the overall group delay function assumes a spiky appearance, due to the group delay taking very large values at sampling points close to the zeros due to the excitation in the z-domain. The strength of these spikes depends upon the proximity of the zeros to the unit circle. Closer the zero to the unit circle, larger is its value. The sign of the spike depends upon whether it lies inside (positive) or outside (negative) the unit circle. The problem in group delay processing of speech signals may thus be posed as one of extracting the characteristics of the system and source from the combined group delay of Fig.2.13 or 2.14.

In the estimation of parameters of the speech signal from the power spectrum of speech we saw that in all the techniques (without exception) an attempt was made (i) to estimate the spectral envelope (for estimating formants) and (i) to flatten the spectrum (for estimating source parameters). This essentially amounts to separating the spectra corresponding to the source and the system.

If the group delay spectrum of speech' is to be used effectively for estimating parameters from the speech signal, the source and

system group delay functions must be necessarily separated. The objective of this thesis is the development of appropriate algorithms for the estimation of the underlying characteristics of the source and system parameters for speech from the group delay spectrum of the given speech signal.

In Chapter 3 we develop a new algorithm for formant extraction from speech using a group delay function derived from the magnitude spectrum of speech. In Chapter 4 we develop yet another algorithm for estimation of formant and pitch from the speech signal using a group delay function (called the modified group delay function) that is derived directly from the phase spectrum.

2.9.2 *Application of Group Delay functions to Spectrum Estimation*

Spectrum estimation is yet another area of signal processing where the dominance of FT magnitude spectrum in most analysis methods is evident. The techniques developed for speech signal do not make any specific assumptions about the signal.

In Spectrum estimation there are two major problems, two types of signals are considered most often, namely, (a) sinusoids in noise and (b) AR processes in noise. The presence of additive noise either introduces new zeros or redistributes the existing zeros. It was shown in the previous section that the zeros due to random noise lie close to the unit circle in the z-domain of the signal. Similarly window zeros lie on the unit circle in the z-domain. If the group delay functions of such signals is to be processed these zeros must be suppressed. The modified group delay function may be thought of as a function in which the information corresponding to that of zeros in a signal are suppressed. The use of modified group delay functions to suppress zeros that are caused by noise in spectrum estimation is studied in Chapter 5.

CHAPTER 3

MINIMUM PHASE CROUP DELAY FUNCTION AND ITS APPLICATION TO FORMANT EXTRACTION FROM SPEECH

3.1 Introduction

In this Chapter we derive a minimum phase signal from the given signal. The group delay function of this signal is then derived which contains information about the location of resonances in the signal. This group delay function is then used to extract formants from speech signals. The algorithm is similar to the cepstral smoothing approach for smoothing the spectrum using homomorphic deconvolution [J.S.Lim; 1979a]. The significant differences are (i) the logarithmic operation is replaced by $(.)^r$ operation and (ii) the additive and high resolution properties of group delay functions are exploited to emphasise formant peaks. The group delay function (or the negative derivative of the Fourier transform phase) is derived for a signal which in turn is derived from the Fourier transform magnitude of the signal. If a suitable value of r is used, this method gives highly consistent estimates of formants compared to both the cepstral approach and the model-based linear prediction (LP) approach for smoothing the magnitude spectrum. The effects of the parameters, exponent r and window width p on the proposed technique of formant extraction are studied.

3.2. Principle of the proposed method

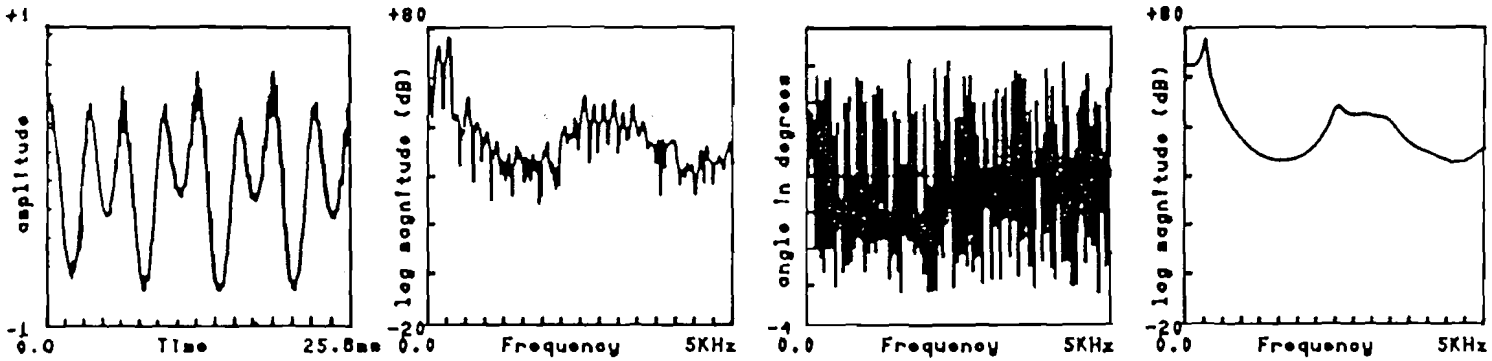
We propose a spectral root group delay function approach for extracting the parameters of the system. This is similar to the spectral root homomorphic deconvolution (SRDS) [J.S.Lim; 1979a]. The proposed method involves deriving a signal with the characteristics of a minimum phase signal so that the phase spectrum of this signal contains the information of the magnitude spectrum. Peaks of the group delay function derived from this phase function correspond to

location of resonances in the signal.

Table.3.1 gives the algorithm for the new spectral root group delay function approach for estimating the group delay function with minimum phase characteristics. In the Table DFT and IDFT correspond to the forward and inverse Fourier transforms, respectively. w(n) is a half Hann window function and is given by

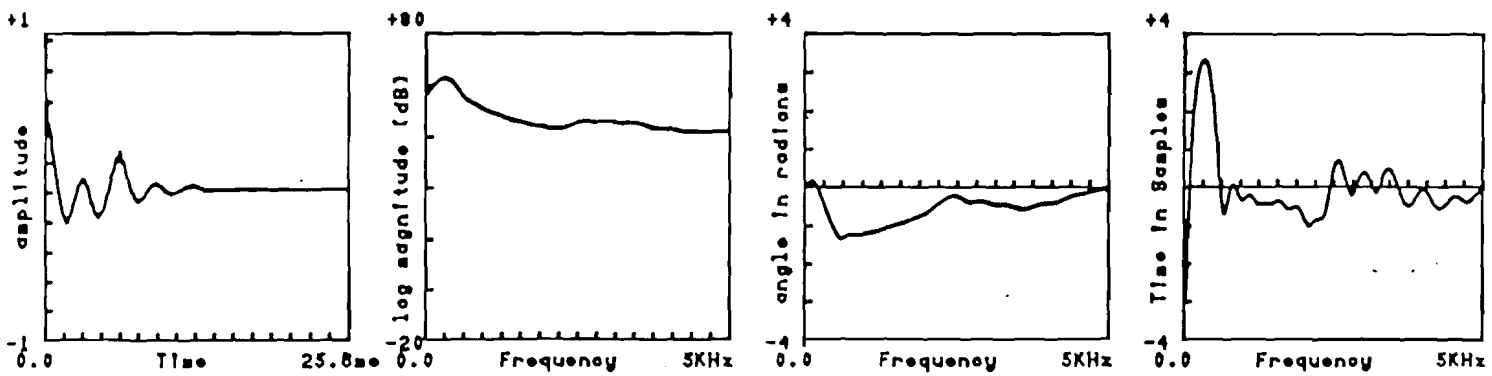
$$w(n) = 0.5 + 0.5\cos(\pi n/L), \quad 0 \leq n \leq L,$$
$$= 0.0, \quad n > L$$

where L is the length of the window. This technique is like the cepstral smoothing technique, except that (i) rth power operation is used in place of the log operation and (ii) the phase group delay is computed instead of the smoothed magnitude spectrum. Fig.3.1a shows a segment of speech (25.6 ms, 10kHz sampling rate). Figs.3.1b,3.1c and 3.1d show the corresponding megnitude, phase and LP spectra. The inverse Fourier transform of the magnitude function gives an even sequence which is called the spectral root cepstrum ($\tilde{x}(n)$). The even sequence is then truncated to include only the causal portion of it. Fig.3.2a shows the causal portion of $\tilde{x}(n)$. The rth power operation does not disturb the locations of either the poles or zeros of the z-transform of the windowed signal. Therefore the region around n=0 in the root cepstrum will contain information corresponding to the slowly varying component of the spectrum and a peak due the periodicity will appear at $n=T_o$, where T_o is the periodicity in the signal. To estimate the parameters due to system this signal is multiplied by a half Hann window to select the first p (corresponding to 4.2ms) samples (henceforth referred as $\tilde{x}_p(n)$). It is worth noting that the magnitude and phase spectra of the original signal are unrelated, whereas the magnitude and phase spectra of the signal $\tilde{x}_p(n)$ are related. Notice that peaks in the magnitude spectra of Fig.3.2b correspond to phase transitions in Fig.3.2c. The



(a) Speech Segment (b) Magnitude Spectrum (c) Phase Spectrum (d) LP Spectrum

Fig.3.1 A segment of speech and its corresponding spectra.



(a) Min. phase signal (b) Magnitude Spectrum (c) Phase Spectrum (d)Group delay Spectrum

Fig.3 Minimum phase signal ant its corresponding spectra

differenced phase corresponds to the group delay function shown in Fig.3.2d. The window size p should be taken as large as possible to obtain a good resolution of formants, but should be less than the periodicity (if it exists) in order to avoid fluctuations due to excitation in the magnitude and phase spectra.

Table.3.1. Algorithm for Computing the minimum phase group delay function from the given signal.

1. Let $x(n)$ be the given N-point sequence. Compute the N-point DFT of $x(n)$, $X(k)$, $k = 0,\ldots,N-1$.

2. Let $\hat{X}(k) = |X(k)|\exp(j\theta(k))$. Compute the N-point IDFT $(|X(k)|^r)$ where r is a value chosen between 0.5 and 1.

3. Let $\tilde{x}(n) = \text{IDFT}(|X(k)|^r)$, $n = 0,\ldots,N-1$. Now multiply $x(n)$ by the window $w(n)$ to eliminate the noncausal portion of the signal and peak due to periodicity.

4. Let $\tilde{x}_P(n) = \tilde{x}(n).w(n)$, $n = 0,\ldots,p$
$= 0$, otherwise.

Compute the N-point DFT of $\tilde{x}_P(n)$, $\tilde{X}_P(k)$, $k = 0,\ldots,N-1$.

5. Let $\tilde{X}_P(k) = |X_P(k)|\exp(j\theta_P(k))$. Compute the group delay function as :

$$\tau(k) = \theta_P(k+1) - \theta_P(k), \quad k = 0,\ldots,N-2$$
$$= \tau(k-1), \quad k = N-1.$$

The strength of our approach is in the fact that, being not model-based, it should provide a better representation of the underlying nature of the system than that obtained using model-based analysis. The cepstral approach to parameter extraction is also not model-based but has the disadvantage that the computation of cepstrum involves a logarithm operation. We now give some of the properties of the spectral root cepstrum.

3.3 *Properties of the spectral root cepstrum*

Let $\{x(n)\}$ be a causal, real and stable sequence and let $\{X(k)\}$ be its discrete Fourier transform.

1. Then $\text{IDFT}(|X(k)|) = \{\tilde{x}(n)\}$ is an even sequence.

2. Given that {x(n)} is a sequence with finite support, from the Akhiezer-Krein and Fejer-Riesz theorems [A.Papoulis; 1977, Ch.7] it can be shown that

$$\begin{aligned} \text{IDFT}(|X(k)|^r) &= \text{IDFT}(|X(k)|^{0.5r}|X(k)|^{0.5r}), \\ &= \text{IDFT}\{Y(k)\}\{Y^c(k)\}, \\ &= \{y(n)\} * \{y(-n)\} \end{aligned} \qquad (3.1)$$

where c and * denote complex conjugation and convolution operations, respectively. Thus $|X(k)|^r$ can be expressed as the Fourier transform of the autocorrelation function of some sequence y(n).

3. Given that $|X(k)|^r$ is a positive even function and that {x(n)} is a non-zero sequence, then $\text{IDFT}(|X(k)|^r)$ is maximum at the origin [R.N.Bracewell;19861.

4. Minimum phase property : The above properties suggest that the truncated sequence $\tilde{x}_p(n)$ behaves like a minimum phase signal in the sense that the phase and the magnitude spectra of $\tilde{x}_p(n)$ are related.

This is confirmed by computing the roots of the z-transform of the sequence $\tilde{x}_p(n)$ numerically for a number of different frames of speech data. It was found without exception that in all the examples ($\approx$ 50) the roots (error < 10^{-12}) lie inside the unit circle (Fig.3.3). We have seen in our studies.that the group delay functions derived from the magnitude and phase of the FT of $\tilde{x}_p(n)$ are identical (Fig.3.4). Also using the properties 2 and 3, the given signal $\tilde{x}_p(n)$ is a signal in which there is minimum energy delay. This is yet another property of minimum phase signals [A.J.Berkhout; 1973,19741.

From these empirical observations we conclude that the causal portion of $\tilde{x}(n)$ can be considered as a minimum phase sequence. The minimum phase condition ensures that the log magnitude and phase of such a signal are related through the Hilbert transform. Thus the complete magnitude information is captured in the FT phase of $\tilde{x}_p(n)$.

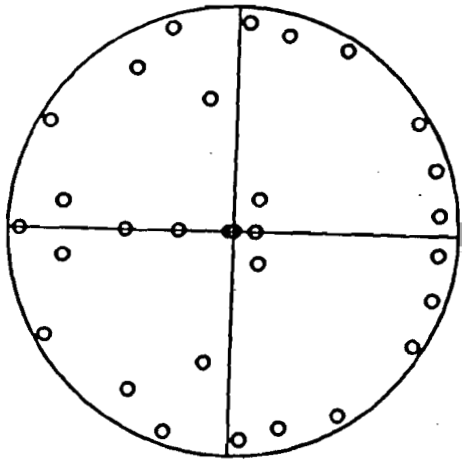
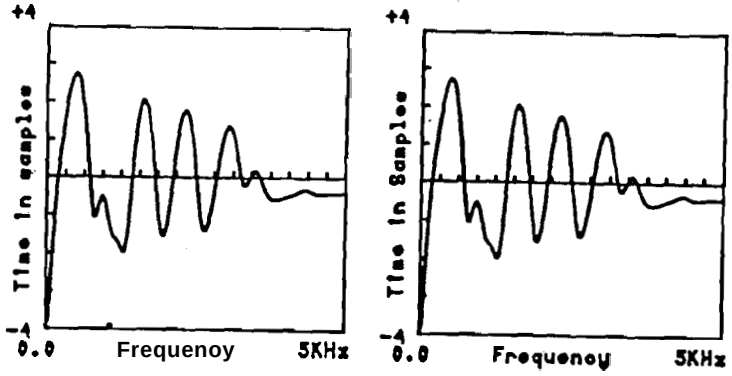


Fig.3.3 Distribution of roots in the z-plane for the minimum phase signal.



(a) Group delay from magnitude

(b) Group delay from phase

Fig.3.4 Illustration of $\tau_p(\omega) = \tau_m(\omega)$ for minimum phase signal.

Because of the minimum phase characteristic, the cepstral coefficients and hence the weighted cepstrum can be derived recursively from $\tilde{x}_p(n)$. The group delay function can also be computed as the FT of the weighted cepstrum [B.Yegnanarayana, D.K.Saikia and T.R.Krishnan; 19841. But in the algorithm discussed in this Chapter we compute the group delay function through the spectrum using the discrete Fourier transform relation.

Note that use of the exponent r on $|X(k)|$ does not alter the peaks in the smoothed envelope of the magnitude spectrum. Thus the peaks locations information is preserved in the computation of the magnitude spectrum of the windowed spectral root cepstrum. It is generally not possible to model each of the peaks by a simple second order all-pole system (resonator) even though they may'correspond to a resonance peak in the original magnitude spectrum of $X(k)$. Therefore low order (10 to 18) linear prediction analysis of the signal $\tilde{x}_p(n)$ will not result in the desired peaks.

It is important to note that the exponent r in $|X(k)|^r$ does not disturb the location of the peak due to periodicity in $\tilde{x}(n)$. Therefore both the spectral peak locations and value of the period are not altered by the exponent factor. But the exponent factor helps to contain the significant information of the spectral envelope in a short window size p for $\tilde{x}(n)$. This helps in the choice of a p lower than that of the periodicity to avoid the influence of periodicity on estimating the system information.

3.4 *Formant Extraction from Speech Using Minimum Phase Croup delay Spectra*

In this section we demonstrate the effectiveness of the proposed group delay function for formant extraction from speech. In the context of speech the peaks of the group delay function correspond to formants. The usefulness of this approach is established by

comparing it with the LP and cepstral approaches for formant extraction. In each case the raw formant data is obtained and the performance is judged by visual inspection of the formant contours. In the group delay function approach for formant extraction there are a few parameters which decide the resolution and accuracy of the formant data that may be obtained. One of them is the window size p. This is similar to the cepstral (or LP) smoothing technique in which the window size (or model order) chosen in the cepstral domain (or LP model) affects the resolution that can be achieved. In addition to the window size p, the exponent r also plays a significant role in the resolution that can be obtained. We now study the effects of varying these two parameters on the formant information obtained from the group delay function. To discuss the performance of our method we first consider synthetic speech data corresponding to the formant contours shown in Fig.3.5.

Model for the synthetic signal :

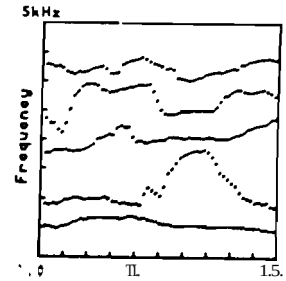
The synthetic signal chosen is a voiced utterance generated by using a simplified model for speech production shown in Fig.3.6. The waveshape for the glottal pulse was chosen to be of the form [L.R.Rabiner and R.W.Schafer; 1978, p.1021:

$$\begin{aligned} g(n) &= 0.5(1 - \cos(\pi n/N_1)), \quad 0 \le n \le N_1 \\ &= \cos(\pi(n - N_1)/2N_2), \quad N_1 \le n \le N_1+N_2 \\ &= 0, \qquad \text{otherwise.} \end{aligned} \tag{3.2}$$

The transfer function for the vocal tract was modelled as :

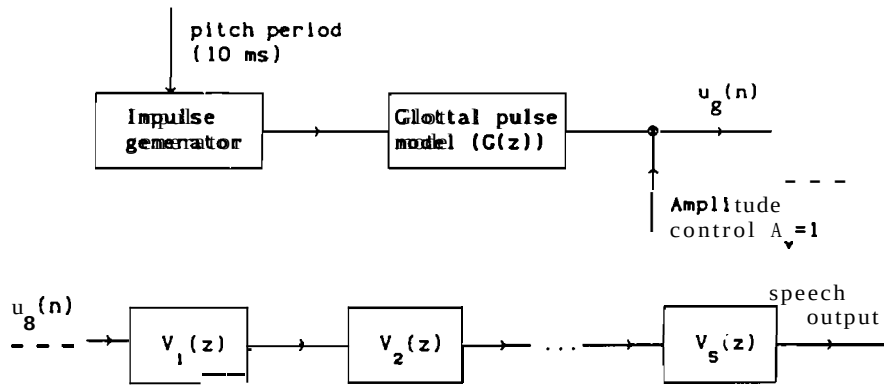
$$V(z) = \prod_{k=1}^{5} \frac{1 - 2e^{-\pi B_k T}\cos(2\pi F_k T) + e^{-2\pi B_k T}}{1 - 2e^{-\pi B_k T}\cos(2\pi F_k T)z^{-1} + e^{-2\pi B_k T}z^{-2}}. \tag{3.3}$$

This equation describes a cascade of digital resonators that have unity gain at zero frequency. All the five formants (F_ks) vary continuously with time as defined by the formant plot shown in

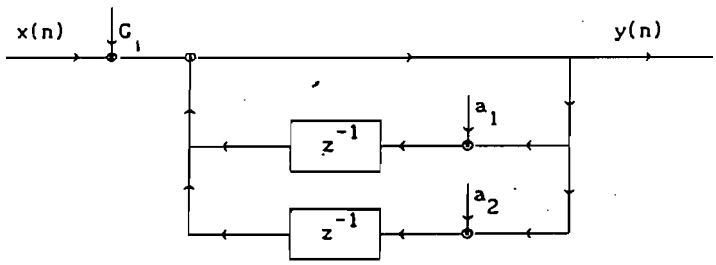


Formant data

Fig.3.5 Synthetic Formant data.



where $V_1(z)$ Is represented by the following filter:



where $G_1 = 1 - 2e^{-nB_1T}\cos(2\pi F_1 T) \cdot e^{-2\pi B_1 T}$

$a_1 = -2e^{-\pi B_1 T}\cos(2\pi F_1 T)$

$a_2 = e^{-2\pi B_1 T}$

Fig.3.6 Model used for speech production in generating synthetic speech.

Fig.3.5. T is fixed at 0.0001sec (i.e. 10kHz sampling rate). The formant bandwidths (B_ks) were fixed apriori at 10% of the formant frequencies.

A pitch period of 10ms was chosen to generate the excitation signal. In the model for the glottal pulse N_1 = 60 and N_2 = 10. The formant data was designed so as to capture most of the situations encountered in practice (in the context of voiced speech), namely, proximity of formants, sudden rise in formants, sudden fall in formants.

The effect of the window size p on formant extraction is studied by obtaining plots of the raw formant data for different values of the window in the spectral root cepstrum domain. Fig.3.7 shows the formant data obtained from the synthetic speech signal using the group delay (GD) approach for various window sizes (**p** = 5.0ms to 8.0ms). The window size is varied uniformly from 5.0ms to 8.0ms in steps of 1.0ms, with **r**=0.5. Notice that an increase in window size results in an increase in resolution of the peaks. The formant data is consistent over a sufficiently large range of window sizes (Fig.3.7a - Fig.3.7c). But too large a window size (for example 8.0ms) causes spurious peaks to appear as in Fig.3.7d. The window size should be large enough to resolve the peaks that are close to each other, but should not be too large to include the effects of pitch on formant extraction.

Fig.3.8 shows the formant data for the same sythetic data using LP analysis for various orders (10 to 22) and Fig.3.9 shows the formant data obtained using cepstrum analysis for the same window widths as used in the GD approach. Comparison of the data in Fig.3.7 with the raw formant data obtained from LP analysis (Fig.3.8) and the raw formant data obtained from cepstrum analysis (Fig.3.9) shows that our method gives equally good but more consistent estimates of

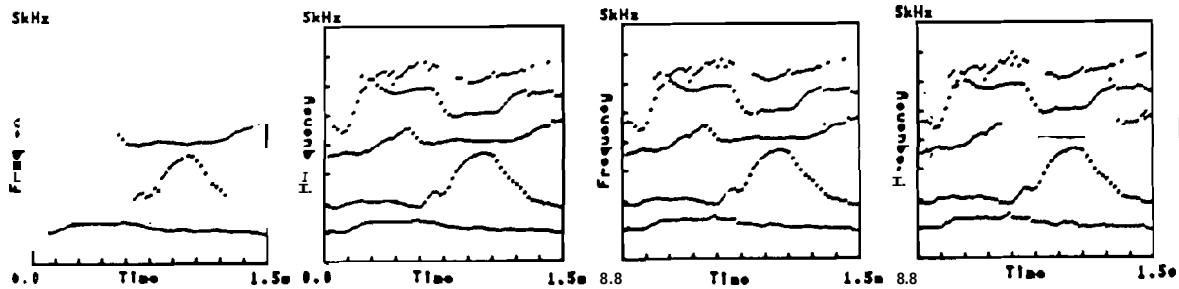


Fig.3.7

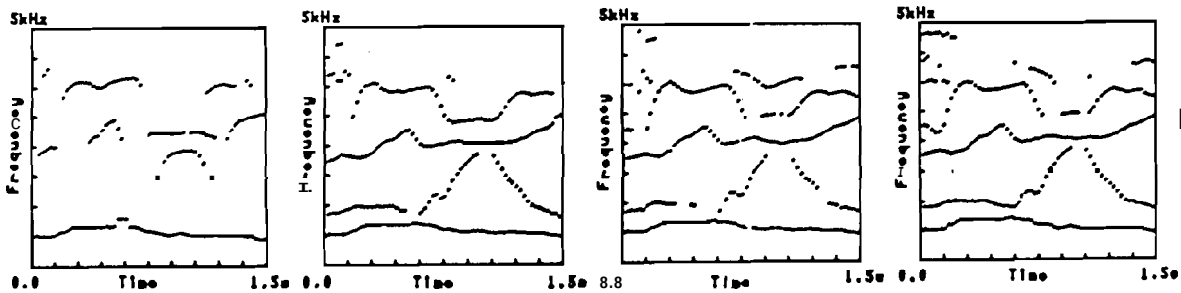


Fig.3.8

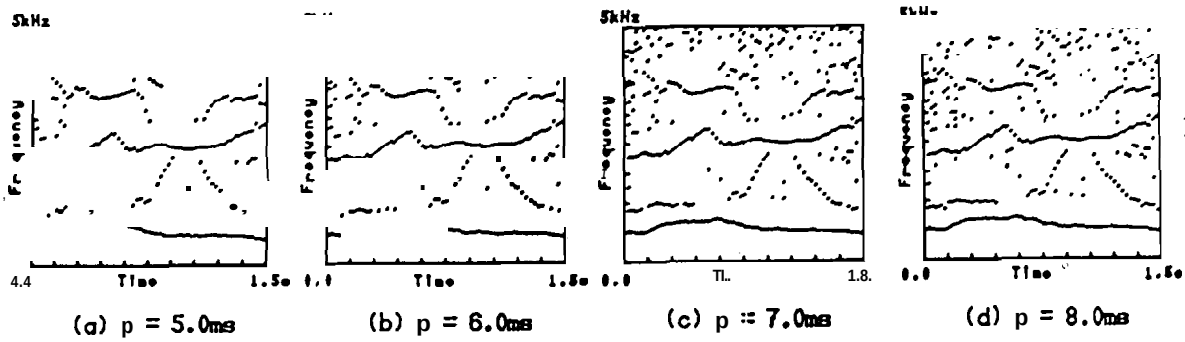


Fig.3.9

Fig.3.7 Raw Formant data obtained using the GD approach for different window sizes (low pitched synthetic speech).

Fig.3.8 Raw Formant data obtained using LP analysis for different model orders (low pitched synthetic speech).

Fig.3.9 Raw Formant data obtained using Cepstrum Analysis for different window sizes (low pitched synthetic speech).

formants over a wide range of window widths for the synthetic data. It is to be noted that in the case of LP analysis (Fig.3.8) lower order may not bring out all the formant peaks. As the order is increased all the peaks get resolved but not many spurious peaks will be generated since the data is strictly the output of an all-pole system.

The choice of **r** (for a particular window size p) depends upon the dynamic range of the signal spectrum. The choice of **r** basically dictates the degree of overlap between the source and the vocal tract components in the root cepstrum domain. In the digital implementation of the SRDS algorithm the $(.)^{r}$ and $(.)^{1/r}$ operations require two phase unwrapping operations [J.S.Lim; 1979a]. In our approach we perform $(.)^{r}$ only on the positive real function $|X(\omega)|$. Hence no phase unwrapping is necessary. Thus there is no constraint on r. We have experimentally observed that the choice of **r** is related to the spectral flatness. It appears logical that when the dynamic range is low (as in the case of noisy speech) the peaks in the spectrum must be emphasized if they have to make a significant contribution to $\tilde{x}(n)$. This can be achieved by keeping $r > 1$. On the other hand, when the dynamic range is very large (as in the case of normal or high pitch voiced speech) the contribution by the first formant dominates the computation of $\tilde{x}(n)$. The effect of the first formant must be deemphasized. This is done by keeping $r < 1$. When $r < 1$ the vocal tract information is concentrated around the origin in $\tilde{x}(n)$ and the gating function enables a good separation of the source information from the vocal tract response. Fig.3.10 illustrates the performance tradeoff for various choices of **r** and window sizes. The effect of **r** can be visualized by traversing Fig.3.10 vertically from bottom to the top along a direction corresponding to a fixed window width.

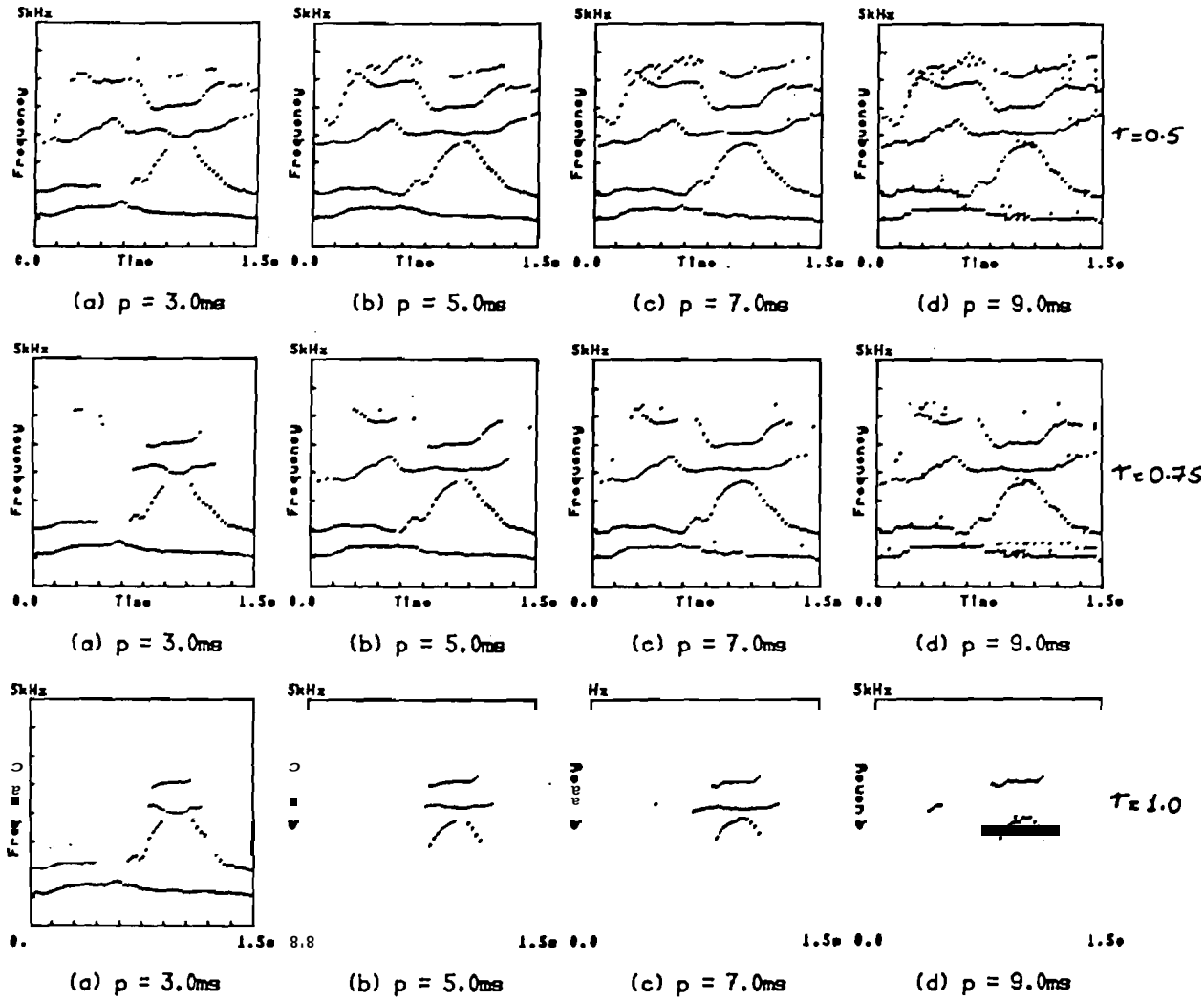


Fig.3.10 Illustration of the GD formant extraction technique for different choices of p and r.

It is to be noted that the parameter r and window size p are related. A smaller window size produces a poorer resolution, while a smaller r produces a better separation of source and excitation. As the window size p must be smaller than the pitch period to avoid fluctuations, r can be manipulated to obtain a good resolution of formants.

So far we have illustrated the use of this new technique for formant extraction on synthetic speech, where the synthetic speech has be modelled as the glottal excitation of a truly all-pole model.

Natural speech may not correspond to a truly autoregressive process of a fixed model order. We now compare the GD approach with that of LP analysis for formant extraction from natural speech. Figs. 3.11, 3.12 and 3.13 shows the formant data obtainkd using GD approach, LP approach and cepstrum analysis for the utterance "We were away a year ago" as spoken by a male speaker. The comparison confirms our earlier conclusions that the GD formant extraction technique gives more consistent formant values (for various window sizes) than that of the LP approach (for various orders) and cepstrum analysis (for various cepstral windows).

Figs.3.14 and 3.15 illustrate the formant contours for a high pitched synthetic and natural speech data. Here r is chosen to be 0.5. As long as the window size p is less than the pitch period, the proposed method works well even for high pitched speech. The synthetic speech was generated using the same procedure indicated in Fig.3.5. The pitch period used for this case was 5ms. For natural speech the utterance is "We were away a year ago" as spoken by a female speaker. In Fig.3.15a, in the region between 0.6 - 0.9s the GD method does not resolve the 2nd and 3rd formants as well as that of the LP method (Fig.3.15b) because the time window chosen is very small (2.4ms). The time window cannot be increased beyond 3.2ms as

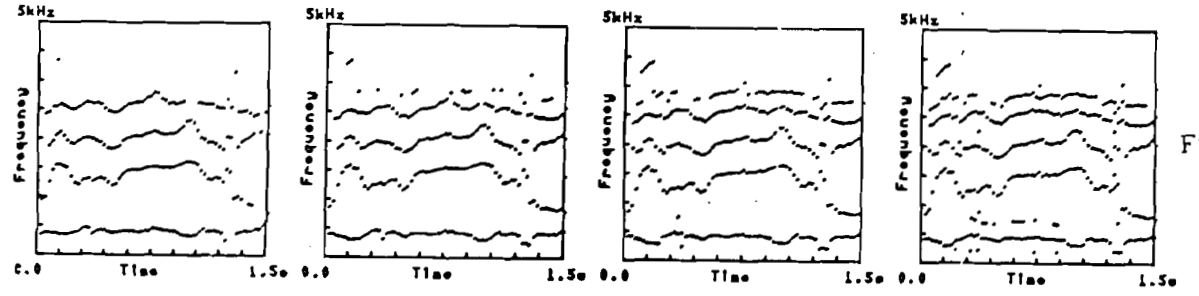


Fig. 3.11

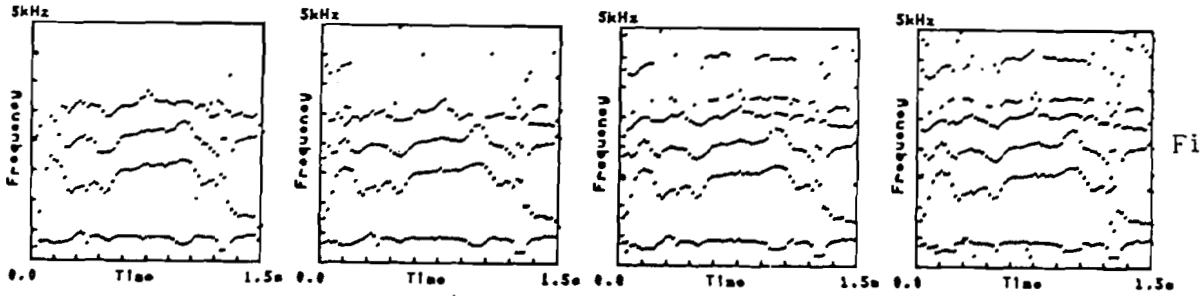


Fig. 3.12

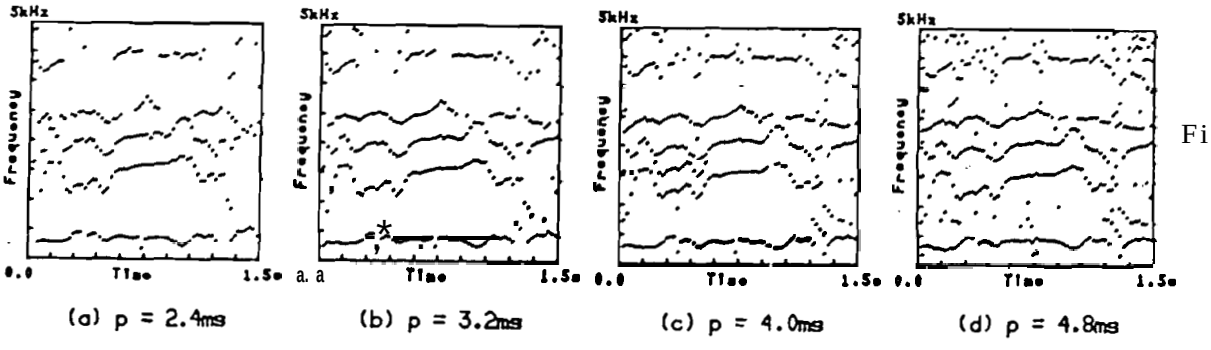


Fig. 3.13

Fig.3.11 Formant extraction Prom natural speech using GD approach (male voice).

Fig.3.12 Formant extraction from natural speech using LP analysis (male voice).

Fig.3.13 Formant extraction from natural speech using Cepstrum analysis (male voice).

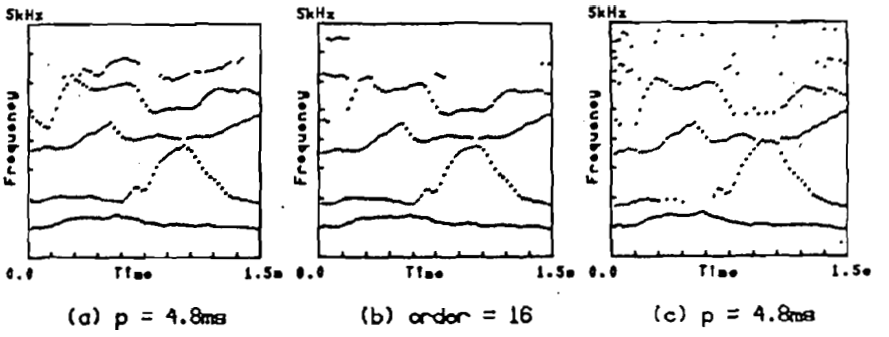


Fig.3.14 Formant extraction from high-pitched synthetic speech using (a) GD approach (b) LP analysis and (c) Cepstrum analysis.

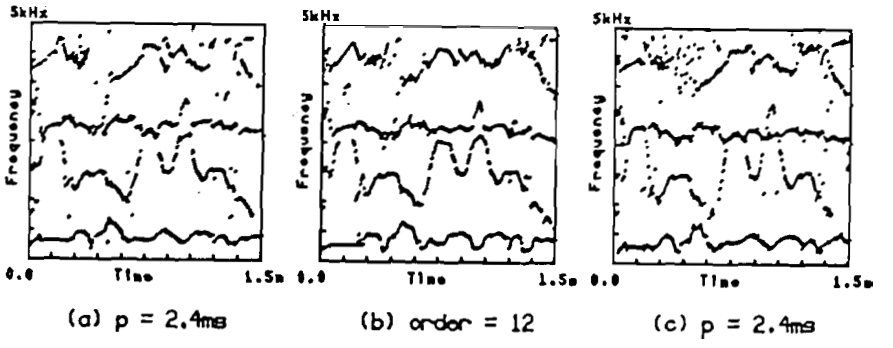


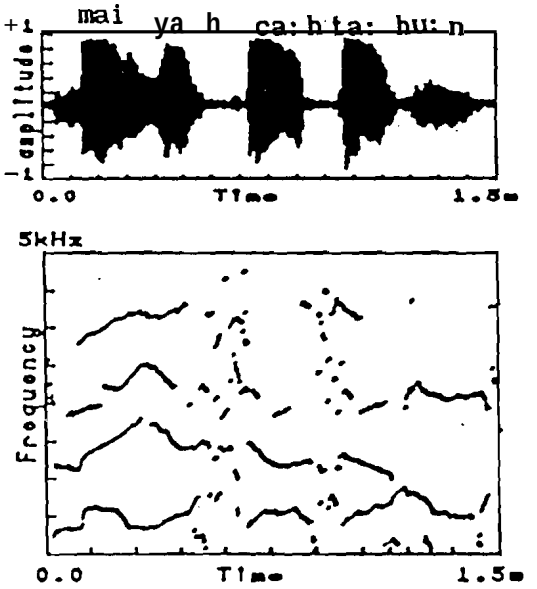
Fig.3.15 Formant extraction for natural speech [female voice) using (a) GD approach (b) LP analysis and (c) Cepstrum analysis.

the average pitch period for this utterance is about 4ms. It is observed that the formants are steady for window sizes ranging from 1.6ms to 3.2ms. For the LP method in Fig.3.15b a carefully chosen order of 12 seems to be appropriate for this utterance. A lower order does not resolve the formants while a higher order generates a lot of spurious peaks. The cepstrum analysis (Fig.3.15c) generates spurious peaks especially at high frequencies for all window sizes.

Fig.3.16 shows the formant contours for an utterance in an Indian language Hindi "mai yah ca:hta: hu:n". The sentence contains segments of different categories of speech segments such as unvoiced, nasals and fricatives. For unvoiced segments the peak locations occur at random frequencies. For most voiced segments the formant frequencies are extracted well as seen from the continuity of the points.

We have examined the performance of the proposed method for noisy speech data also. Fig.3.17c shows the formant contours obtained for an utterance (Fig.3.17a) with an overall SNR = 10 dB, using p = 3.2ms and r = 2. The variation of SNR for each frame is shown in Fig.3.17b. The formant contour for the clean data is given in Fig.3.17d. Comparison of Figs3.17c and 3.17d show that there are spurious peaks at those frames where the SNR is very low (< 0 dB), while for all other frames the formant peaks even for the noisy data are located at the appropriate frequencies. For some segments in the noisy data in the region 0.9s to 1.2s (in Fig.3.17c), the fourth formant is not extracted as well as that for the clean data. This is because for a given frame SNR is a function of frequency also. At high frequencies usually the SNR is lower than that at low frequencies.

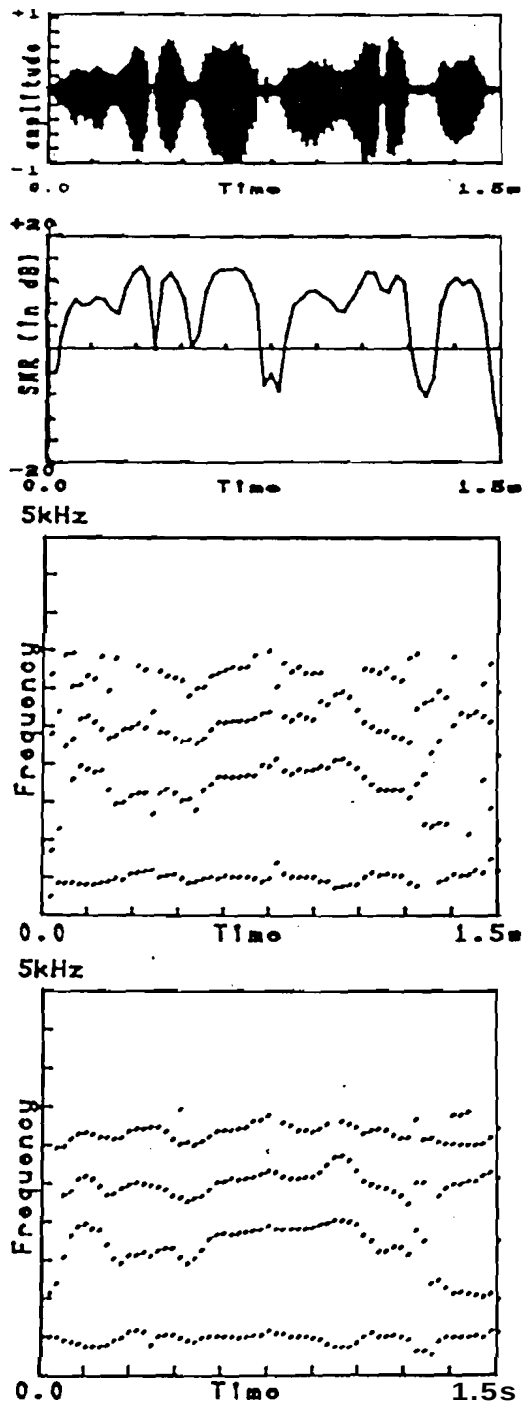
While the proposed method seems to work well for a wide variety of speech signals, computation time is significantly higher than the



(a) Speech Signal

(b) Formant Data

Fig.3.16 Formant extraction for an utterance in an Indian language Hlndl containing different categories of speech segments including nasals and unvoiced.



(a) Speech Signal

(b)SNR as a function of time

(c) Formant Data(Noisy)

(dl Formant Data(Clean)

Fig.3.17 Formant extraction from noisy speech

LP or cepstral methods. For the utterance "We were away a year ago" the computation time for the proposed method is 40 sec, whereas it is 20 sec fcr LP and 30 sec for the cepstrum method on a microvax Vaxstation.

3.5 Summary

We have proposed a new method of extracting formant information from the speech signal. We have demonstrated that the additive and high resolution properties of the group delay functions can be used for extracting closely spaced and low amplitude formant information. This method for formant extraction gives a more consistent performance compared to other methods based upon smoothing the magnitude spectrum. This method of formant extraction does not depend on any model, hence the formant information obtained should be a better representation of the underlying nature of the signal than that obtained from model-based techniques. These studies show that there is a relationship between the spectral flatness and analysis parameters, which can be exploited to choose appropriate values of p and r.

CHAPTER 4

MODIFIED GROUP DELAY FUNCTIONS AND ITS APPLICATION TO SPEECH ANALYSIS

4.1 Introduction

In this Chapter we propose new methods of processing group delay functions to estimate parameters corresponding to the system and source in a source-system model for signal production, where the system corresponds to that of an all-pole system and the source may be a train of impulses or random noise. Generally the estimation of periodicity corresponding to the excitation and the resonances corresponding to that of the system are treated as two distinct problems. In the methods presented in this Chapter, estimation of both system and and source parameters involve similar analysis methods. Therefore, both the problems are addressed in this Chapter.

In the previous Chapter, to overcome the problem of wrapping of phase, a mimimum phase group delay function was derived from the magnitude spectrum to estimate the vocal tract parameters. In this Chapter we suggest a new method of processing the group delay function directly. The expression for the computation of the group delay function is modified to derive a modified group delay (MGD) function. In the MGD function, the large amplitude spikes due to the source are suppressed. Parameters corresponding to the system, namely, frequencies of resonances are extracted from this modified group delay function.

Periodicity in a signal manifests as a sinusoidal component in the spectrum. A modified group delay function for this sinusoidal component is obtained. Peaks appear at regular intervals in the modified group delay function. The distance between two peaks

measured in seconds corresponds to the periodicity.

An approximate isolation of source and system characteristics in the modified group delay function is possible because of the distinct characteristics of the group delay functions of the source and system. In Section **4.2** we study the properties of group delay functions for speech-like signals. The basis for the proposed method - modified group delay functions - is discussed in Section **4.3.** Extraction of both system and source parameters is also discussed in Section **4.3.** In Section **4.4** the performance of this method for different choices of parameters is discussed. The modified group delay function has some interesting properties which make it a good tool for processing noisy speech. In Section **4.5** we address the problem of extracting formants and pitch from both clean and noisy speech. We also discuss a method of synthesising speech from formant and pitch data.

4.2 *Theory* and *Properties 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 visualise 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

$$H(\omega) = \prod_{i=1}^{M} \frac{1}{(\alpha_i^2 + \beta_i^2 - \omega^2 - 2j\omega\alpha_i)}, \qquad (4.1)$$

where $(\alpha_i \pm j\beta_i)$ is the complex pair of poles of the **ith** resonator.

The magnitude spectrum is given by

$$|H(\omega)|^2 = \prod_{i=1}^{M} \frac{1}{[(\alpha_i^2 + \beta_i^2 - \omega^2)^2 + 4\omega^2\alpha_i^2]}, \qquad (4.2)$$

and the phase spectrum is given by

$$\theta(\omega) = \angle H(\omega) = \sum_{i=1}^{M} \tan^{-1} \frac{2\omega\alpha_i}{\alpha_i^2 + \beta_i^2 - \omega^2} \qquad (4.3)$$

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)

$$\tau(\omega) = -\frac{d\theta(\omega)}{d\omega} = \sum_{i=1}^{M} \frac{2\alpha_i(\alpha_i^2 + \beta_i^2 + \omega^2)}{(\alpha_i^2 + \beta_i^2 - \omega^2)^2 + 4\omega^2\alpha_i^2} \qquad (4.4)$$

It was shown in [B. Yegnanarayana; 1978] 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.4.1d. Fig.4.1a shows the windowed impulse response of a 10th order all-pole filter. Figs4.1b, 4.1c and 4.1d show the corresponding magnitude, phase and group delay spectra. Note that the group delay function (Fig.4.1d) 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.

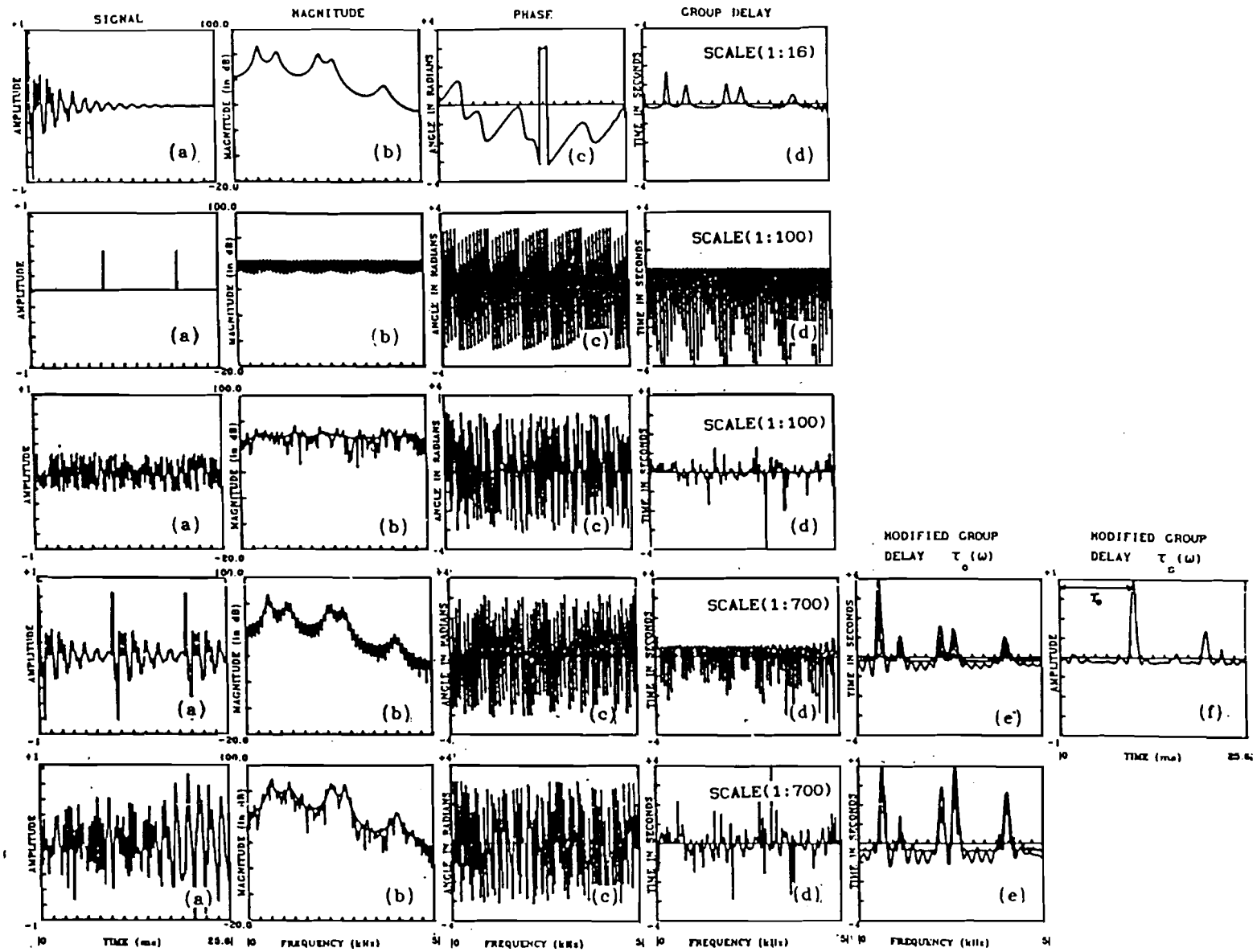
Fig. 4.1 Impulse response of a 10th order all-pole filter and its corresponding spectra.

Fig. 4.2 Impulse train and its corresponding spectra.

Fig. 4.3 Random noise and its corresponding spectra.

Fig. 4.4 Response of all-pole filter to impulse train and its spectra.

Fig. 4.5 Response of all-pole filter to random noise and its corresponding spectra.



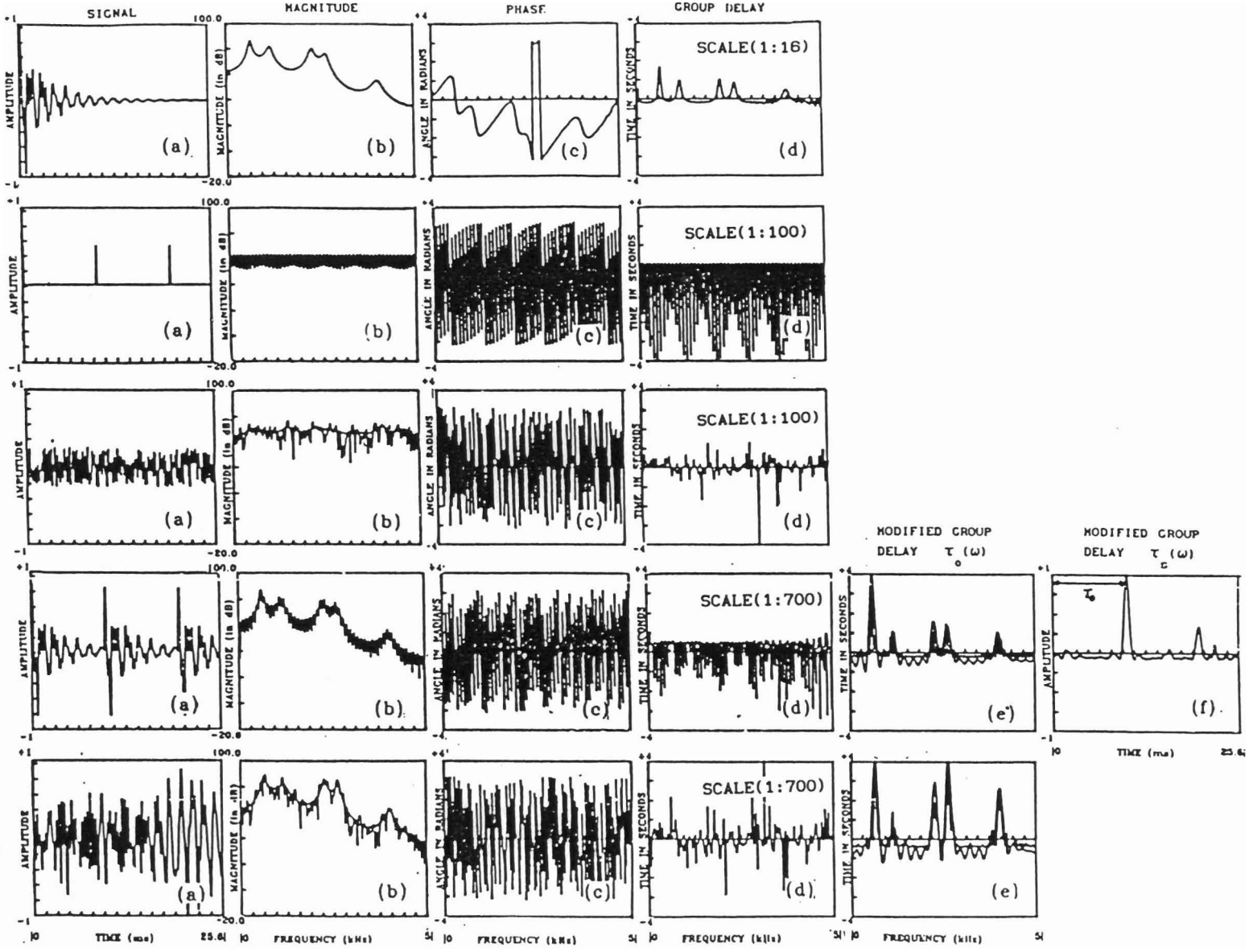
F g. 4. 1 Impulse response of a 1 th order all-pole filter and i s corresponding spectra.

Fig. 4.2 Impulse train and its corresponding spectra.

Fig. 4.3 Random noise and its corresponding spectra.

Fig. 4.4 Response of all-pole filter to impulse train and its spectra.

Fig. 4.5 Response of all-pole filter to random noise and its corresponding spectra.



It was shown in [K.V.Madhu Murthy and B.Yegnanarayana; 1989] 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. Therefore if the group delay function of a signal is to processed, the signal should be sampled adequately. The sampling frequency required being dictated by the locations of the zeros of the signal z-transform. In a source-system model for signal production (all-pole system and all-zero source) to estimate the parameters of the model corresponding to that of the system, if the group delay function corresponding to that of the source is eliminated in the overall group delay function, then sampling based on the Nyquist criterion will result in proper sampling in the group delay domain. This is precisely what is attempted in the technique proposed in this Chapter.

4.3 *Basis for the Proposed Method : Hodified Group Delay functions*

4.3.1 Extraction of System parameters

In digital processing of signals like speech, the vocal tract system and the excitation contribute to the envelope and the fine structure, respectively to 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 system characteristics from the group delay function, the component due to spectral fine structure must be deemphasised. These spikes form a significant part of the fine structure and their effect cannot be eliminated by normal smoothing techniques.

In our previous attempts [Hema A.Murthy, K.V.Madhu Murthy and

B. Yegnanarayana; 1989a], the signal was modified prior to the group delay computation to reduce the effect of the spikes in the group delay domain. In this technique an attempt is made to smooth the phase spectrum and then compute the group delay function. This is done by taking the average of complex spectral values at three points. $\omega-\delta\omega$, ω and $\omega+\delta\omega$. The phase of the averaged spectrum is first computed. The group delay function corresponding to this phase spectrum is then computed. It was observed that the technique worked quite well provided a zero is located at ω and there are no zeros at $\omega\pm\delta\omega$. It also required a different choice of $\delta\omega$ for different locations of zeros of the signal z-transform. For practical signals like speech, the value of $\delta\omega$ cannot be decided apriori as the locations of zeros in the signal z-transform are determined by both the analysis window and source excitation.

We now suggest a method for reducing the contribution of the fine structure to the group delay function by modifying the expression for computing 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 zeros close to the unit circle. Our initial attempts to compensate for the zeros involved modifying the expression for computing the group delay function in an adhoc manner which was reported in [Hema A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana; 1989b]. We now substantiate this conjecture with both a theoretical analysis and experimental results and suggest a modification which does not involve empirical choice of parameters.

Any signal can be characterised as the response of an all-pole filter to an all-zero excitation, the z-transform of the system generating the signal can be written as

$$H(z) = \frac{N(z)}{D(z)}. \qquad (4.5)$$

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 system. The frequency response of H(z) is given by

$$H(\omega) = \frac{N(\omega)}{D(\omega)}, \qquad (4.6)$$

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

$$\tau(\omega) = \tau_N(\omega) - \tau_D(\omega), \qquad (4.7)$$

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 equation (4.4) earlier. Although it is difficult to derive an analytical expression for $\tau_N(\omega)$, we can study its behaviour in terms of the characteristics of the excitation signal. Since N(z) corresponds to the z-transform of the excitation signal, the zeros 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 zeros are lying inside or outside the unit circle in the z-plane. Figs.4.2 and 4.3 illustrate the behaviour of $\tau_N(\omega)$ for a random noise sequence and impulse train respectively. Note that the log magnitude spectra (Figs.4.2b and 4.3b) have nearly a flat spectral envelope with rapid fluctuations superimposed on it due to zeros close to the unit circle. The group delay function (Figs.4.2d and 4.3d) has large random 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.4.4 and 4.5. The signal in Fig.4.4 corresponds to a windowed version of the signal generated by convolving the impulse train (Fig.4.2a) with the impulse response (Fig.4.1a) of an all-pole system. The group delay function (Fig.4.4d), which is simply the sum of the plots of Fig.4.1d and 4.2d shows that the resonance peaks are indeed masked by the large amplitude spikes. Note that the vertical scales'in Figs4.1d and 4.2d are different, the peak amplitudes in Fig.4.2d being very much larger than the amplitudes in Fig.4.1d. Similar behaviour is observed in Fig.4.5, where the signal is a windowed version of the signal obtained by convolving the random noise in Fig.4.3a with the impulse response (Fig.4.1a) of an all-pole system.

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

$$\tau(\omega) = \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{|X(\omega)|^2}, \qquad (4.8)$$

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 zeros 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

$$\begin{aligned}\tau(\omega) &= \tau_N(\omega) - \tau_D(\omega) \\ &= \frac{\alpha_N(\omega)}{|N(\omega)|^2} - \frac{\alpha_D(\omega)}{|D(\omega)|^2},\end{aligned} \qquad (4.9)$$

where $\alpha_N(\omega)$ and $\alpha_D(\omega)$ are the numerator terms of (4.8) for $\tau_N(\omega)$ and $\tau_D(\omega)$, respectively.

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

$$\tau_m(\omega) = \tau(\omega)\,|N(\omega)|^2$$

$$= \alpha_N(\omega) - \frac{\alpha_D(\omega)}{|D(\omega)|^2}|N(\omega)|^2 \qquad (4.10)$$

In equation (4.10) the contribution of the first term $a_N(\omega)$ should be small compared to the second term in order to emphasise the group delay component of the second term. The analytical proof for the case where the excitation is a train of two impulses separated by a period T_o and the system is a single resonance is given in Appendix B. Fig.4.6a shows the group delay function of an all-pole filter with a single resonance at ω_o. Fig.4.6b shows the zero spectrum corresponding to two impulses. Fig.4.6c shows the plot of the group delay function obtained by exciting the said all-pole filter with the impulses. Fig.4.6d shows the corresponding modified group delay function. Notice that the overall peak of the modified group delay function coincides with that of the group delay function corresponding to that of Fig.4.6a.

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 zeros so that it cancels the small values in the denominator of the first term in (4.9). Therefore $|N(\omega)|^2$ should retain all the sharp fluctuations

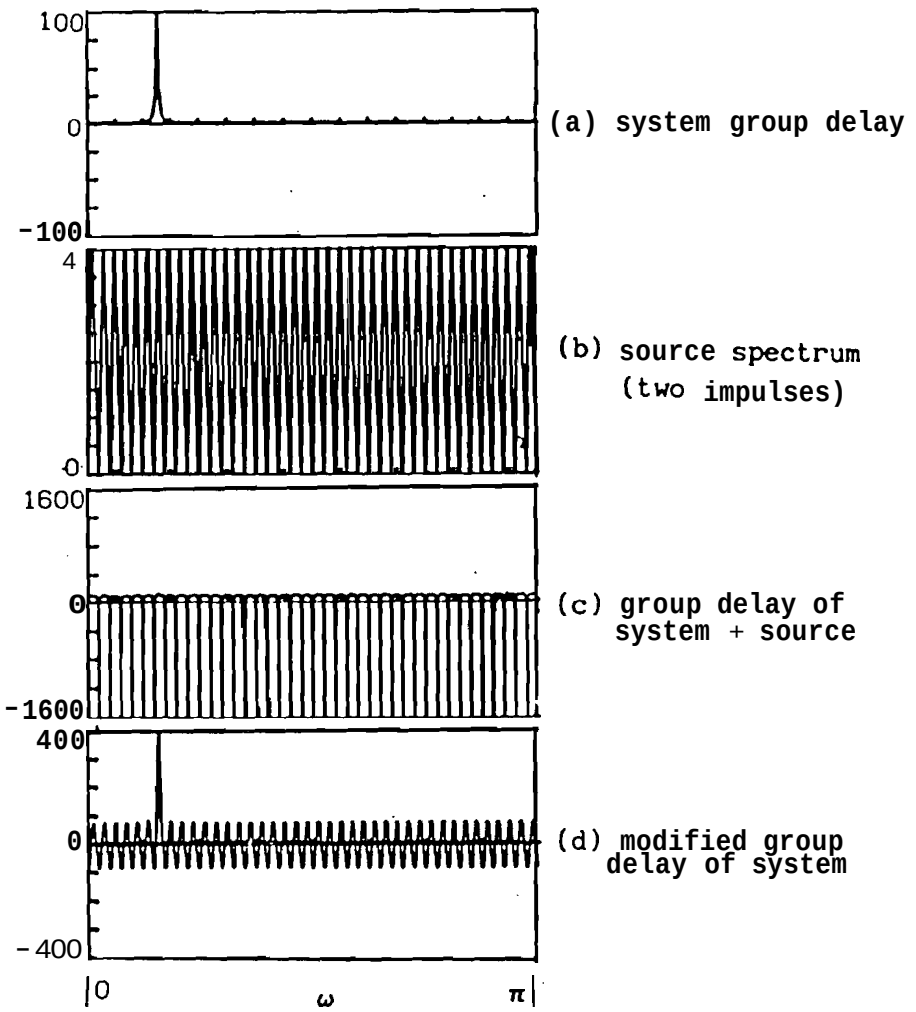


Fig.4.6 Estimation of the modified group delay function corresponding to that of the system in a source-system model for signal production.

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 $V_c(\omega)$ [L.R.Rabiner and R.W.Schafer; p.519, 1978]. That is

$$\hat{Z}(\omega) = \frac{S(\omega)}{V_c(\omega)}, \qquad (4.11)$$

where $S(\omega)$ is the signal spectrum and $V(\omega)$ is the cepstrally smoothed spectrum of $S(\omega)$. Figs.4.4e and 4.5e 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.4.1d. Table.4.1 gives the algorithm for computing the modified group delay function for a given sequence x(n). Alternatively, τ_o can be computed by modifying equation (4.8) for the computation of the group delay as

$$\tau_o(\omega) = \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{V(\omega)}, \qquad (4.12)$$

where $V(\omega)$ is the cepstrally smoothed spectrum of $S(\omega)$.

That $a_N(\omega)$ is small for an impulse train can be seen from Fig.4.4e, where the modified group delay function $\tau_o(\omega)$ is plotted for the signal in Fig.4.4a. 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.4.4e and 4.1d. Note that the modified group delay function in Fig.4.4e is obtained by multiplying the function in Fig.4.4d with an

Table.4.1 Algorithm for Computing the modified group delay function from the given signal.

1. Let x(n) be the given M-pt causal sequence. Compute y(n) = nx(n).
2. Compute the N-pt (N >> M) discrete Fourier transform (DFT) X(k) and Y(k) of the sequences x(n) and y(n) respectively, k = 0,1,...,N-1.
3. Compute cepstrally smmothed spectrum V(k) of $|X(k)|^2$.
4. Compute the modified group delay function τ (k) as

$$\tau(k) = \frac{X_R(k)Y_R(k) + X_I(k)Y_I(k)}{V(k)}, \qquad k=0,I,\ldots,N-1.$$

where R and I denote the real and imaginary parts respectively.

estimate of the excitation spectrum in Fig.4.2a. Fig.4.5 illustrates similar results for the random noise excitation. Later in the experiments we show that for a variety of excitation functions $a_N(\omega)$ is small.

However, it is important to note that the location of a zero (due to the excitation) in the z-domain must not coincide with the location of a pole corresponding to that of a resonance. It will not be possible to suppress the information corresponding to that of source using the modified group delay.

4.3.2 Extraction of Source parameters

In this Section we show that the characteristics of the modified group delay function discussed in the previous Section can be used to derive the periodicity in the excitation signal. Assume that the excitation is periodic with some period T_o. Let us consider the z-transform of two impulse separated by T_o. Then

$$E(z) = 1 + z^{-T_o}, \tag{4.13}$$

$$|E(\omega)|^2 = 2 + 2\cos\omega T_o. \tag{4.14}$$

In the frequency domain $|E(\omega)|^2$ has a periodic component with period $1/T_o$ (pitch frequency). If a zero spectrum, corresponding to the FT magnitude spectrum with a flat spectral envelope, is derived for a voiced speech segment, then the resulting signal contains a sinusoidal component with period $1/T_o$. We now replace ω by n and T_o by ω_o and remove the dc component to obtain a signal

$$s(n) = \cos n\omega_o, \quad n = 0,1,\ldots,N-1. \qquad (4.15)$$

The z-transform of this signal is given by

$$S(z) = \frac{1 - 2\cos\omega_o(N-1)z^{-N} + z^{-2N}}{1 - 2\cos\omega\, z^{-1} + z^{-2}} \qquad (4.16)$$

We use the technique described in Table.4.1 to derive the modified group delay function ($\tau(\omega)$) corresponding to this signal s(n). The numerator polynomial of (4.16) corresponds to the zeros due to the finite window applied in the time domain. The argument used in the previous Section applies for the suppression of window zeros also Fig.4.4f shows the modified group delay function $\tau_s(\omega)$.

4.4 Effects 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 [K.V. Madhu Murthy and B. Yegnanarayana; 19891 and our own experience with the use of group delay functions over the past several years.

Composite signals of the form shown in equation (4.17) below are used in these experiments. Each signal is obtained as the response of a cascade of five resonances to a train of impulses separated by a period p. The amplitude of the impulses are 1, γ, γ^2, γ^3, ... The composite signal is given by

$$y(n) = x(n)+\gamma x(n - p)+\gamma^2 x(n - 2p)+\gamma^3 x(n - 3p)+\ldots \quad (4.17)$$

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

$$Y(z) = \frac{X(z)}{1 - \gamma z^{-p}} \quad (4.18)$$

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 distributon of zeros and the number of zeros are determined by the values γ and p respectively. If $\gamma = 0$, we only have the basic signal. In the following experiments a particular parameter is varied, the modified group delay functions, $\tau(\omega)$ corresponding to the vocal tract system and $\tau(\omega)$ corresponding to the source are computed. The performance is judged by comparing (i) $\tau_o(\omega)$ with $\tau(\omega)$ for the system for synthetic signals and (ii) $\tau(\omega)$ and the time domain signal for source information.

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 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 nature of the phase data due to wrapping. This wrapping problem is absent in the group delay function plot as the group delay function is computed directly from the time domain signal. However, the group delay function appears to be featureless due to effect of zeros close to the unit circle. In the modified group delay plots, the features

corresponding to system and source are emphasised. In all the figures vertical scale is not explicitly mentioned, since we are only looking at the features in the plots.

Experiment **No.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 eq.(4.11) or eq.(4.12) to derive $V_c(\omega)$.

(b) Size and shape of the analysis window for the signal.

(c) Proximity of zeros to the unit circle by varying γ in eq.(4.17).

(d) Number of zeros by varying p in eq.(4.17).

(e) Proximity of resonances.

Experiment **No.1a** : We have found that the modified group delay function is almost the same over a range 4 to 20 samples of cepstral window used to derive $V_c(\omega)$ in eq.(4.12) (Fig.4.7).

Experiment **No.1b** : Effect of windows : For this experiment we set γ = 1 in equation (4.17). The plots for four different windows are shown in Fig.4.8. The interesting part of these results is that the fluctuations caused by zeros due to windows are practically eliminated in $\tau_o(\omega)$. However, the window effect is reflected in the bandwidth of the resonances of the vocal tract system as seen from $\tau_o(\omega)$ corresponding to the system. Note that the resonances are sharp for the rectangular window compared to that for the Hann window. This is a significant result because one of the most important problems in signal processing is to overcome the ripple effects in the spectrum caused by the window in the time domain. The windows do not seem to affect the estimation of the source parameters as can be seen from the $\tau_s(\omega)$ corresponding to the source excitation.

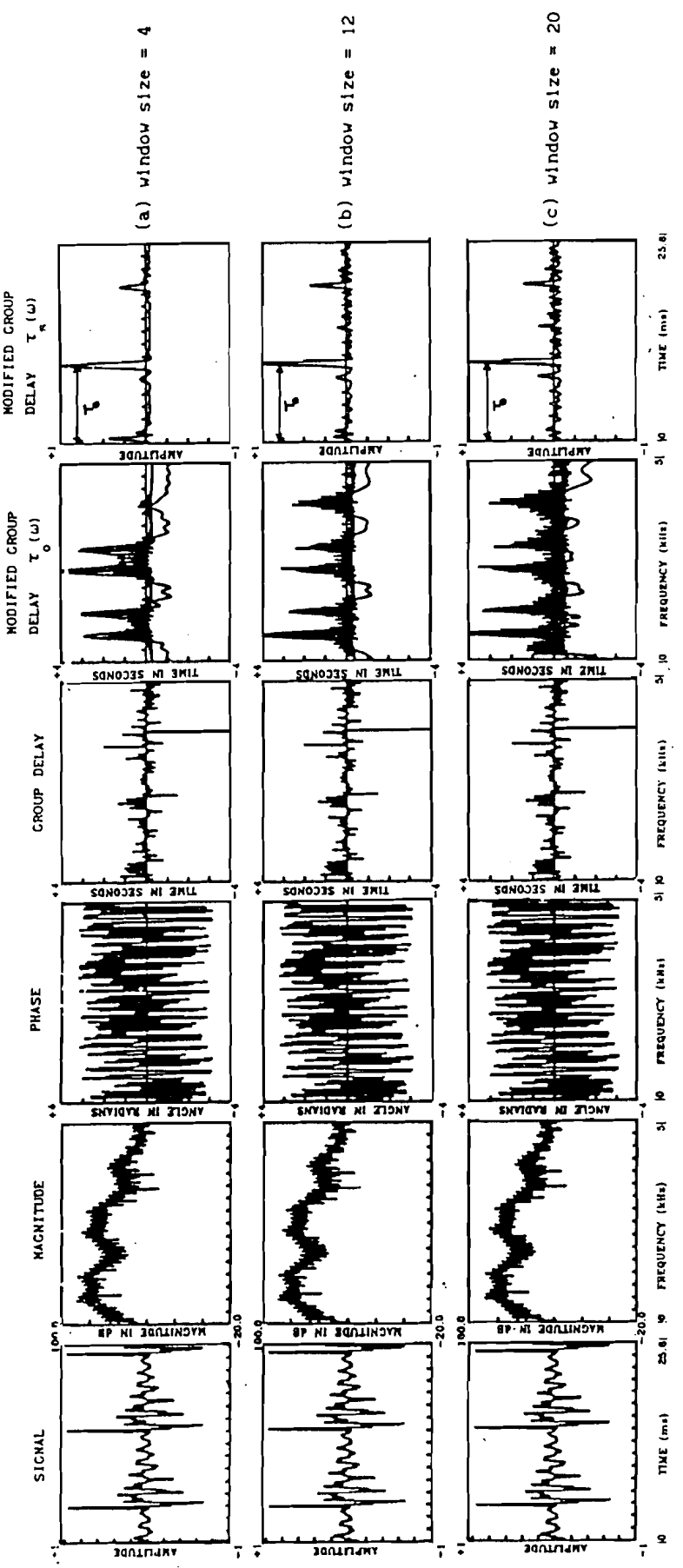


Fig.4.7 Illustration of the effect of different cepstral windows on the modified group delay functions.

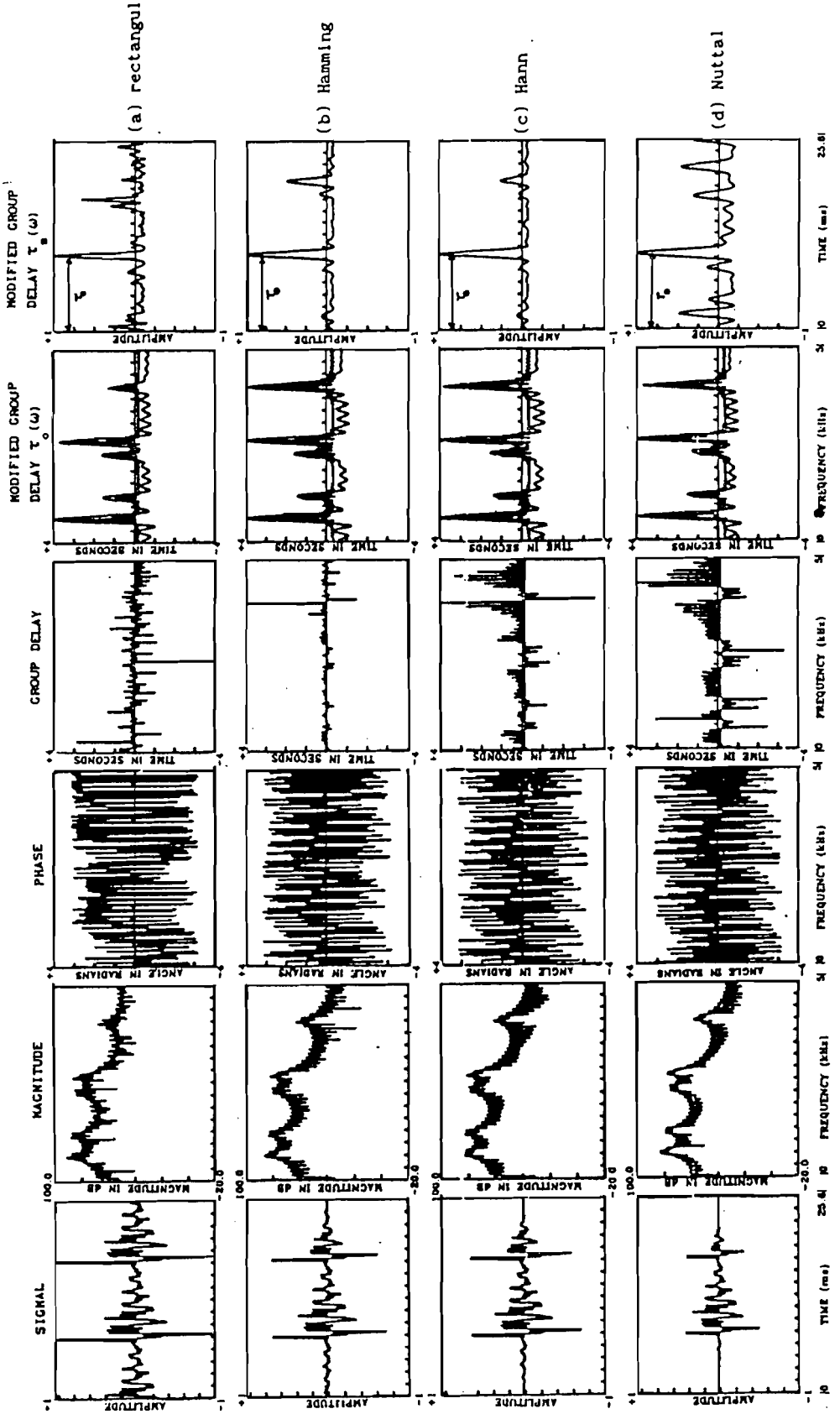


Fig.4.8 Illustration of the effect of different data windows on the modified group delay functions.

Experiment **No.1c** : Effect of varying the proximity of zeros to the unit circle : From equation (4.17) it is seen that by changing the value of γ, which is the ratio of the amplitudes of two successive impulses, we can move the zeros along a radial line in the z-plane. An experiment was conducted in which γ was varied from 0.75 to 1.25. Fig.4.9 shows the modified group delay functions $\tau_m(\omega)$ and $\tau_s(\omega)$, for different values of γ. Notice that the location of the resonances and periodicity are clearly visible in the modified group delay functions $\tau_o(\omega)$ and $\tau_s(\omega)$.

Experiment **No.1d** : Effect of number of zeros : To study the effect of number of zeros on the estimation of system and source information we performed the following experiments. The delay p determines the number of zeros in the z-plane. Thus by varying p the number of zeros in the z-plane can be varied. The value of p is varied from 50-130. Fig.4.10 shows the results. Notice that the effect of the number of zeros on the group delay function is considerably reduced in the modified group delay function of $\tau_o(\omega)$ while $\tau_s(\omega)$ is not significantly altered.

Experiment **No.1e** : Proximity of resonances (Resolution properties) : In this experiment the resonances F_2 (2nd) and F_3 (3rd) are brought close to each other. The difference between the resonances is reduced from 500 Hz to 100 Hz (Fig.4.11). In all cases the resonances are resolved in the modified group delay function. 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 Fourier transform of the given data for computation of the modified group delay function. $\tau_s(\omega)$ is not plotted here as this experiment is not relevant for estimating the periodicity in the signal.

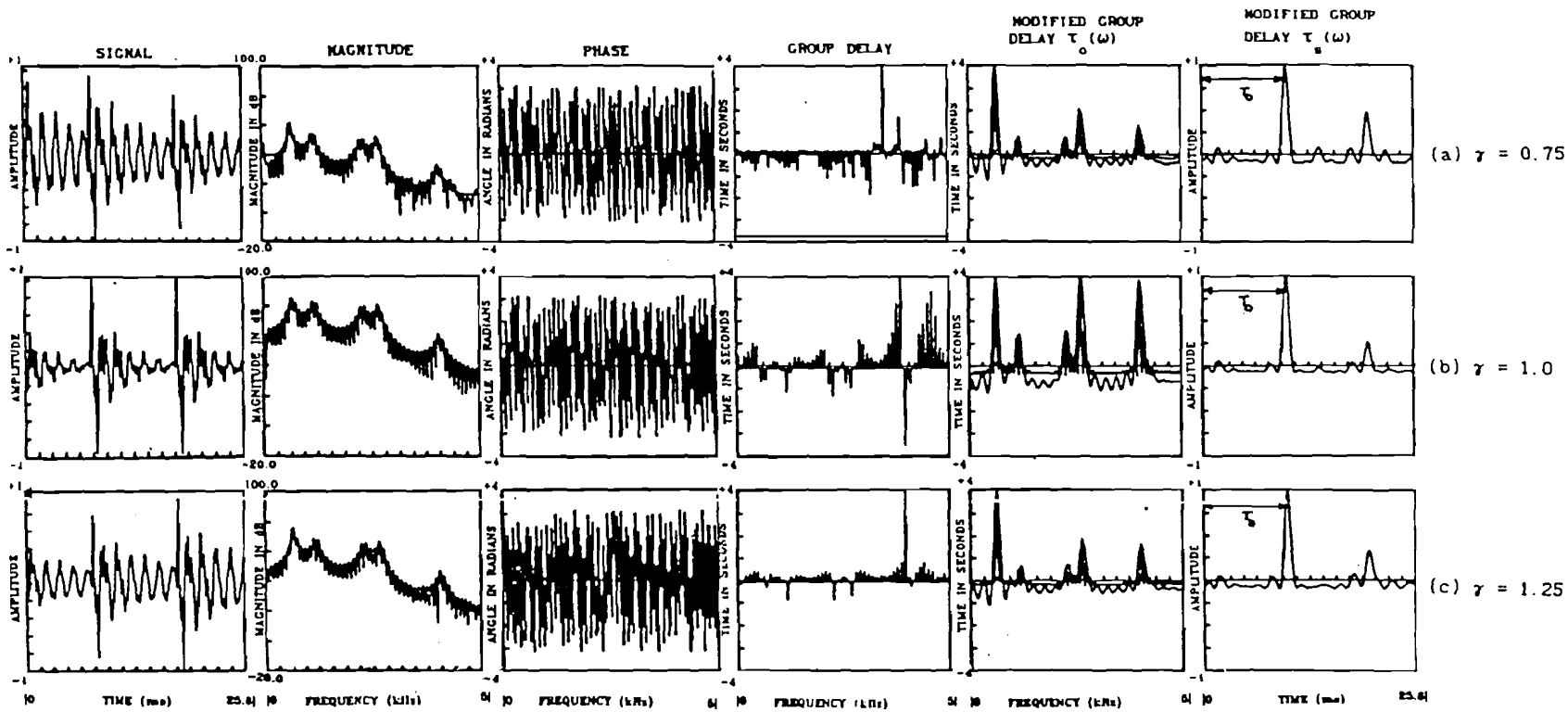


Fig.4.9 Illustration of the effect of varying the proximity of zeros to the unit circle on modified group delay functions

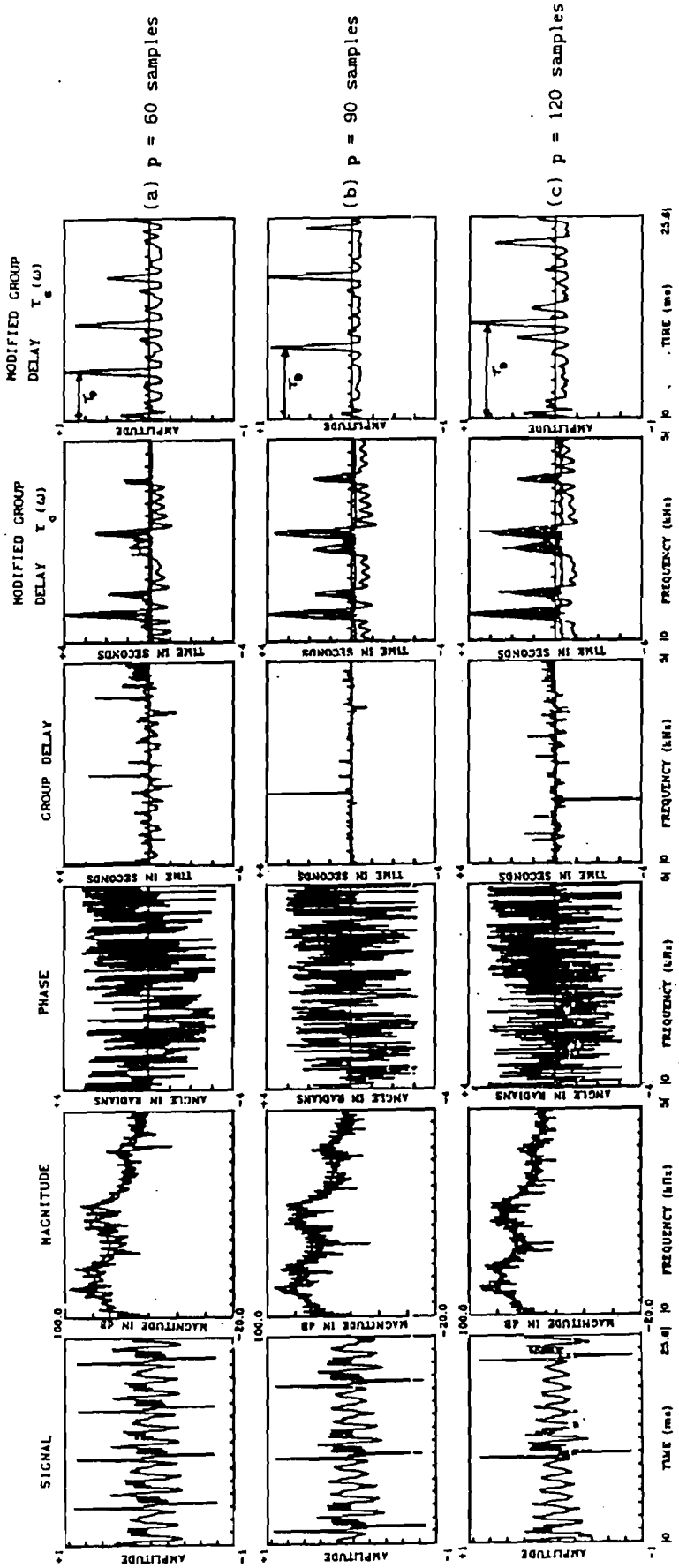


Fig.4.10 Illustration of the effect of number of zeros on the modified group delay functions.

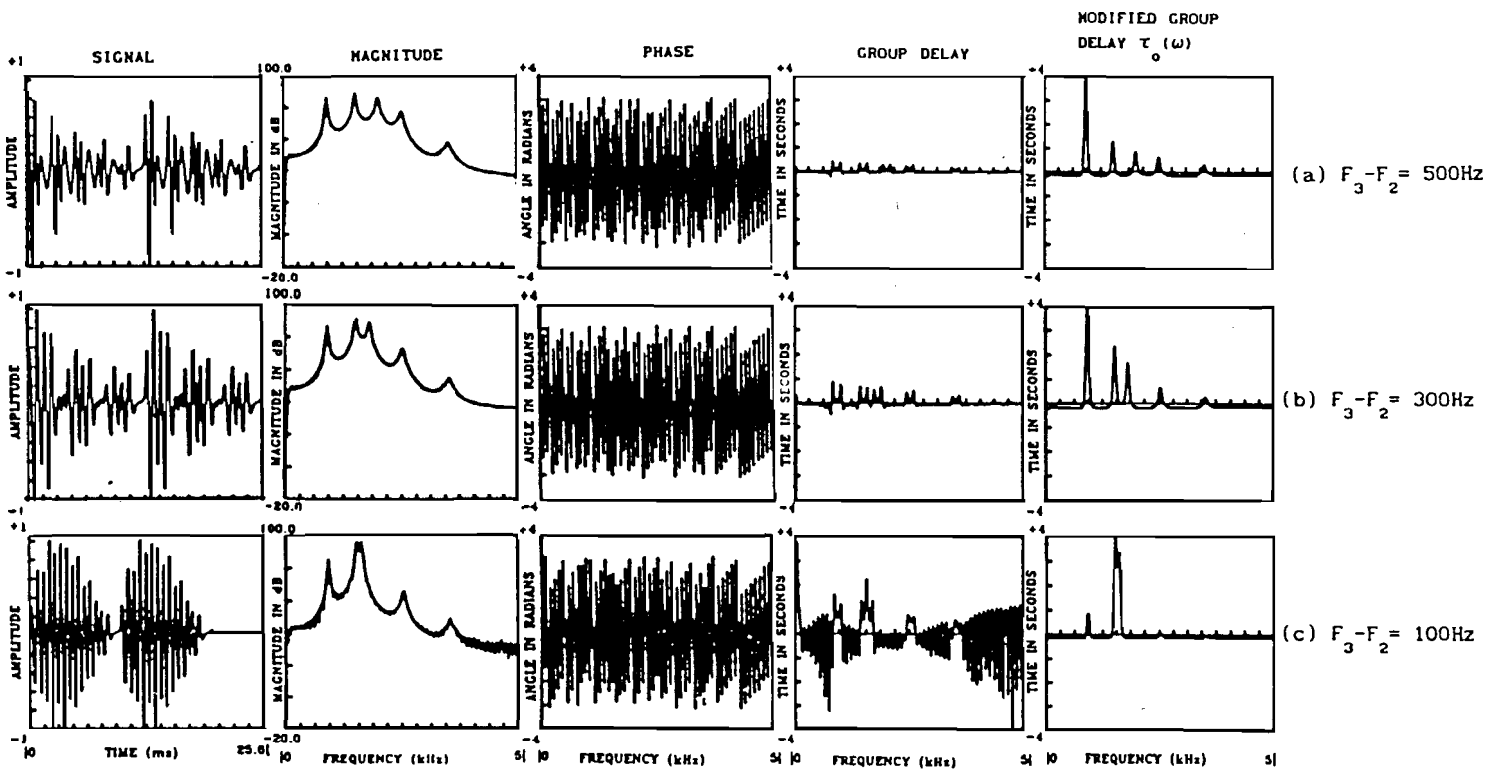


Fig.4.11 Illustration of varying the proximity of resonance on modified group delay functions.

Experiment **No.2:** Different Types of Excitation Functions

So far we have considered 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 four different excitation functions : (a) An impulse sequence separated by a period (100 samples) (b) Synthetic glottal pulse sequence as defined by Rosenberg [L.R.Rabiner and R.W.Schafer, p.103. 19781 and (c) Glottal pulse sequence with radiation load [L.R.Rabiner and R.W.Schafer, p.102, 19781 and (d) Uniformly distributed random noise. The choice of these excitation functions is based upon the model signals used for the excitation in the speech production mechanism. Fig.4.12 shows the results for the different excitation functions. We can see that the effect of these excitation functions on the modified group delay functions is minimal. This is due to the fact that all the excitation functions are finite duration signals which introduce zeros in the z-plane.

Experiment No.3: Natural Speech

In this experiment we consider different segments of natural speech. Fig.4.13 shows the plots for four consecutive segments of speech chosen arbitrarily from an all voiced utterance. The results show that the formant and pitch information are preserved in the modified group delay functions.

Experiment **No.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.4.14. Notice that significant features are preserved even when the SNR is 0 dB. This point is also illustrated in Fig.4.15 for natural speech.

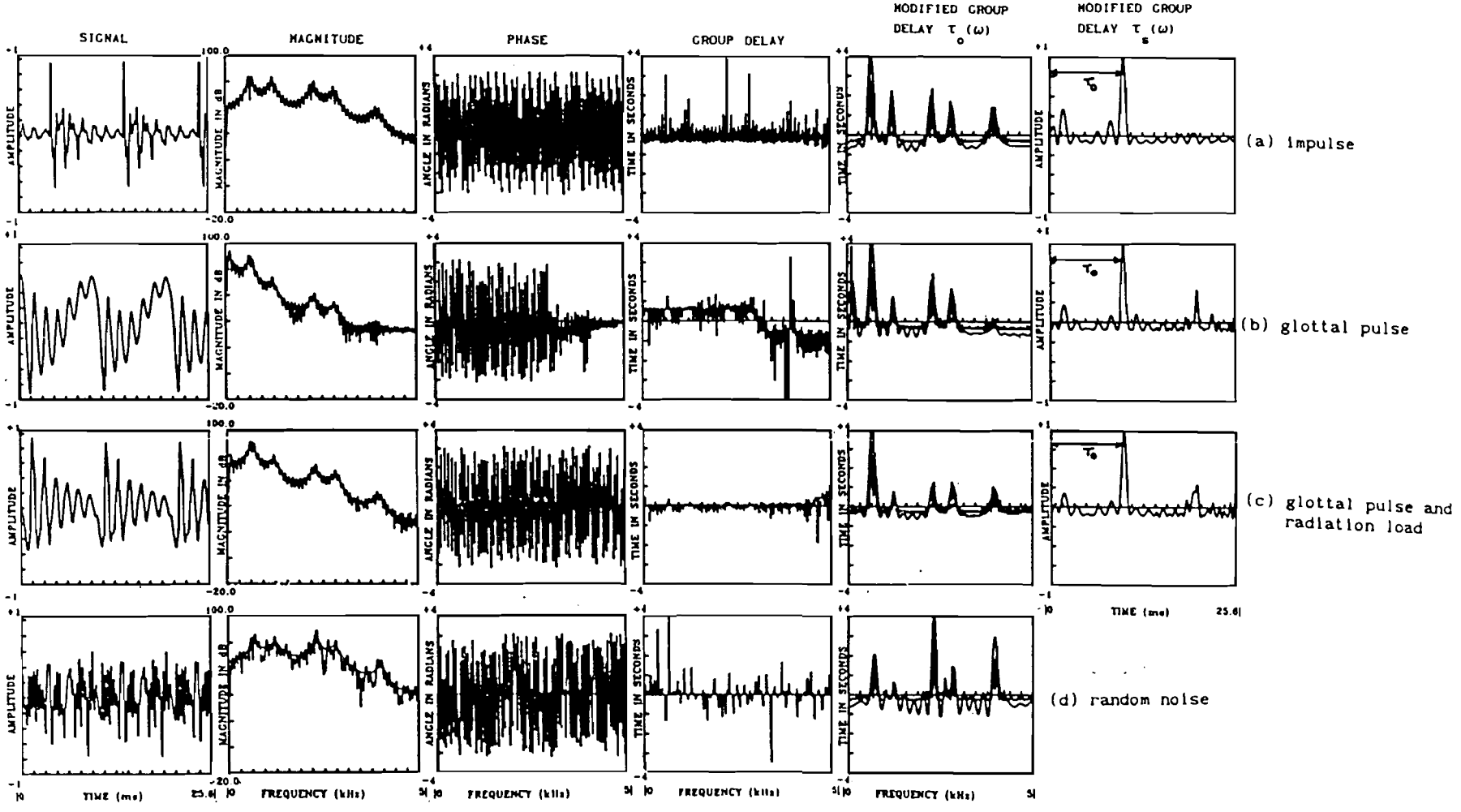


Fig.4.12 Illustration of the effect of different excitation functions on modified group delay functions.

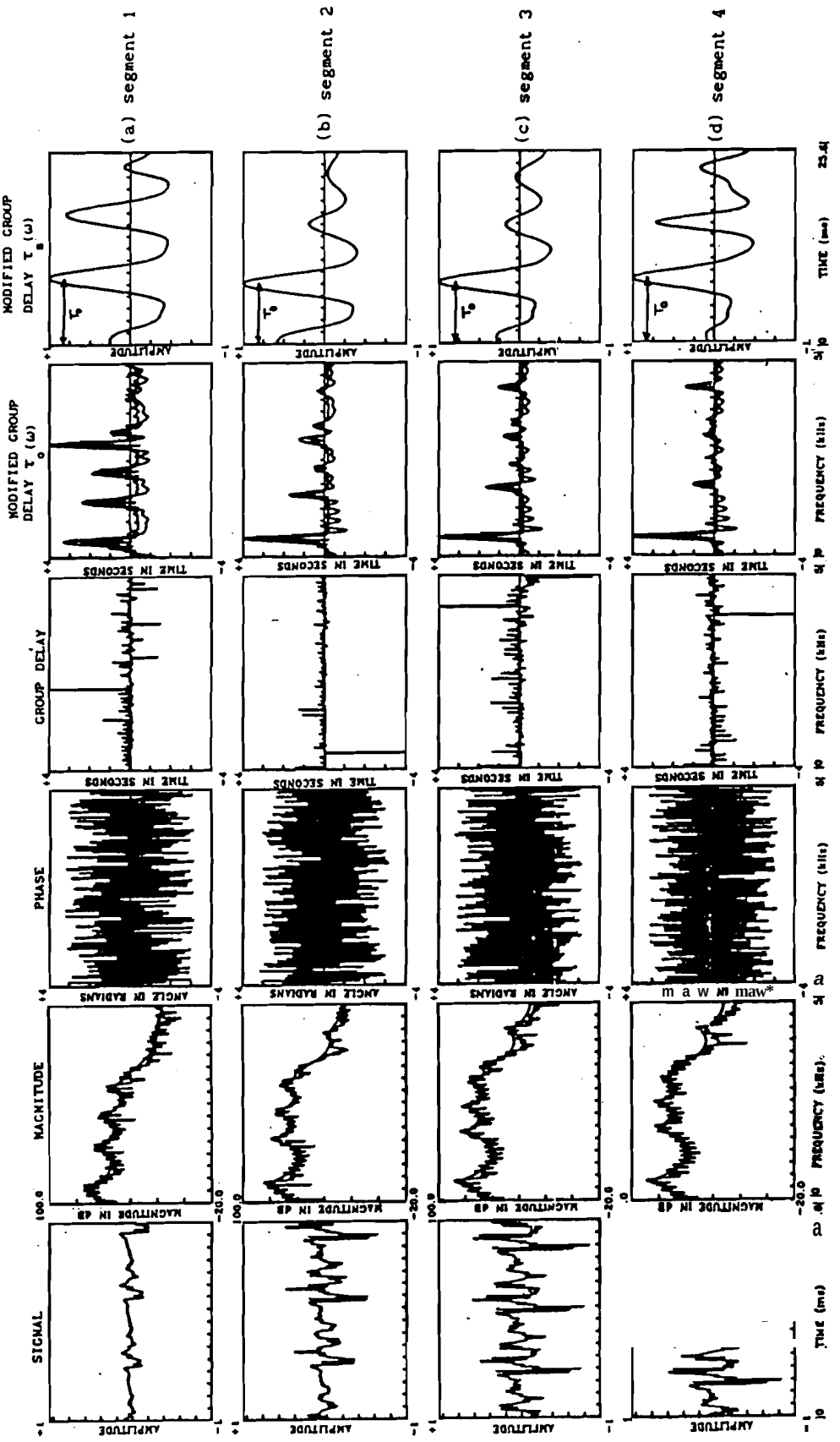


Fig.4.13 Illustration of modified group delay functions for different segments of natural speech.

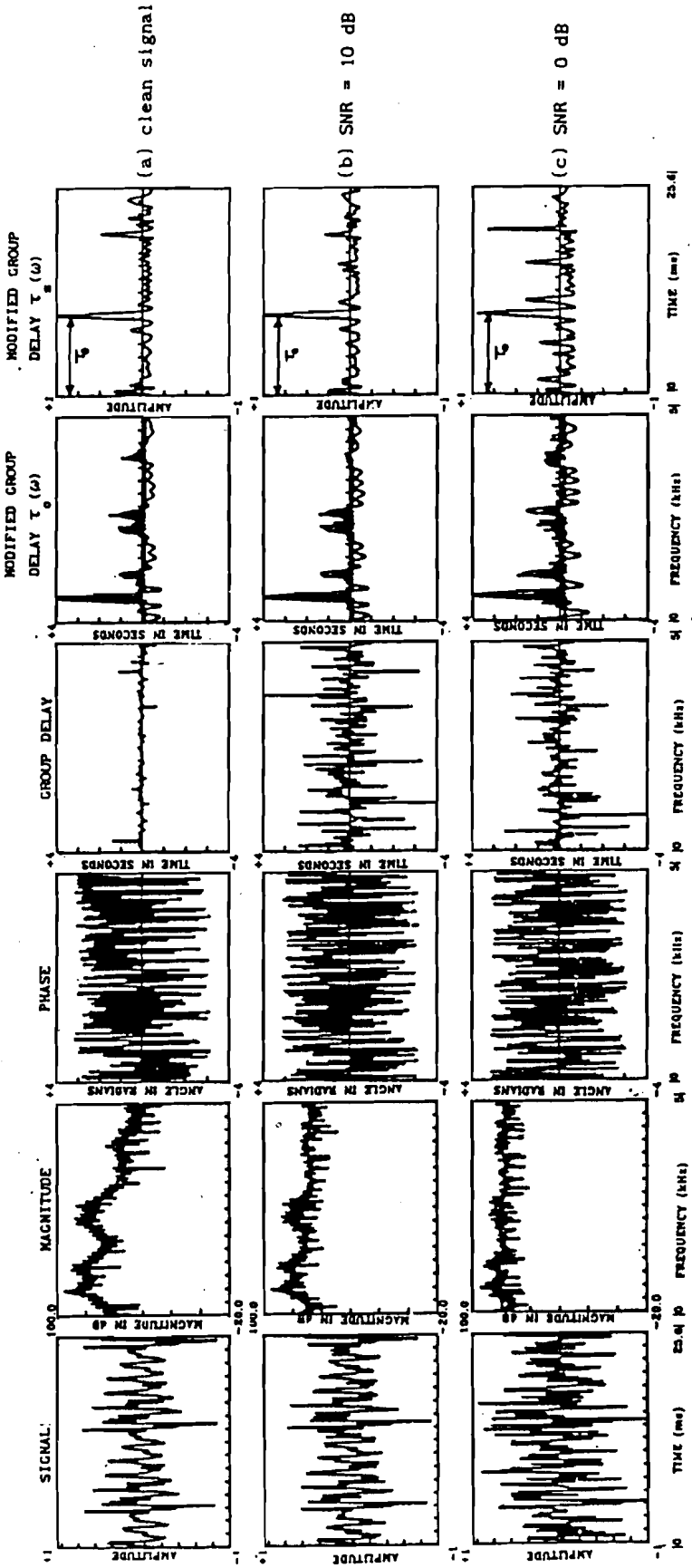


Fig.4.14 Effect of noise on modified group delay functions (synthetic speech).

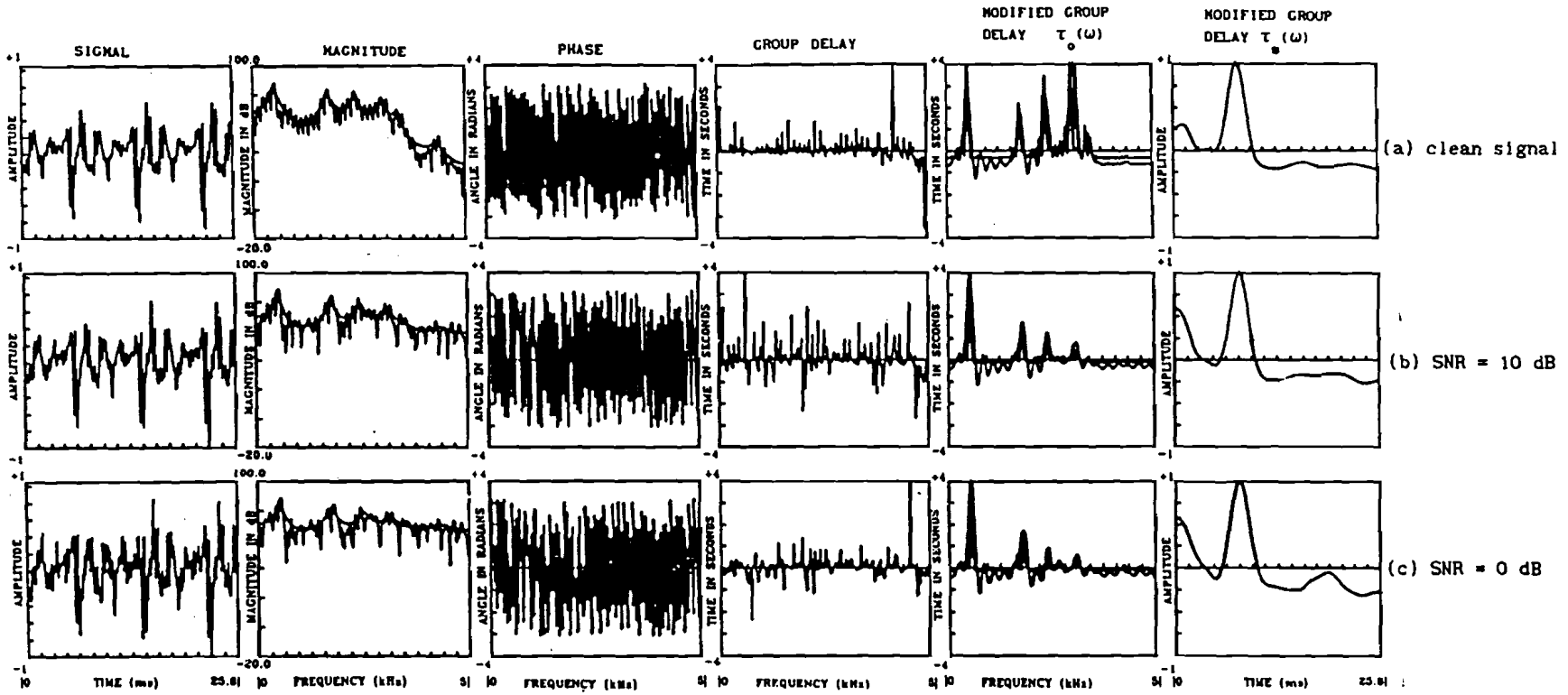


Fig.4.15 Effect of noise on modified group delay functions (natural speech).

Experiment **No.5:** Formant and Pitch Extraction from Speech:

From the various experiments done so far we can conclude that the modified group delay functions derived for the source and system can be successfully used to estimate parameters of model corresponding to the system and source from natural signals like speech. For speech signals the peaks of the smoothed MGD (system) should correspond to formants and the distance of the first peak from the origin measured in seconds of the MGD (source) should correspond to pitch. Fig.4.16 shows an utterance "We were away a year ago" as spoken by a male speaker and the corresponding pitch and formant data obtained using modified group delay functions.

4.5 Speech Enhancement Using Modified group Delay functions

4.5.1 Estimation of Parameters from Noisy Speech

In this section we develop the theory and discuss a technique for enhancing the characteristics of the vocal tract system from noisy speech. The characteristic we are looking for are the formants(resonances) of the vocal tract system and the pitch period of the glottal excitation. We ignore for the time being the effects of data windows.

We define our problem as follows:

Given a noisy signal

$$x(n) = e(n)*h(n) + u(n) \qquad (4.19)$$

where $h(n)$ is the impulse response of the all-pole system $G/A(z)$ and $e(n)$ is either a periodic train of pulses or random noise sequence, determine the resonances of the all-pole system and periodicities of the excitation signal.

Equation (4.19) can be expressed in terms of z-transform as

$$X(z) = E(z)H(z) + U(z) \qquad (4.20)$$

$$H(z) = G/A(z) \qquad (4.21)$$

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \qquad (4.22)$$

The frequency response is given by

$$X(\omega) = V(\omega)/A(\omega) \qquad (4.23)$$

$$V(\omega) = GE(\omega) + A(\omega)U(\omega) \qquad (4.24)$$

The group delay function is defined as the negative derivative of the Fourier transform(FT) phase of a signal. Let $\tau_X(\omega)$, $\tau_V(\omega)$ and $\tau_A(\omega)$ represent the group delay functions corresponding to $X(\omega)$, $V(\omega)$ and $A(\omega)$, respectively. Then

$$\tau_X(\omega) = \tau_V(\omega) - \tau_A(\omega) \qquad (4.25)$$

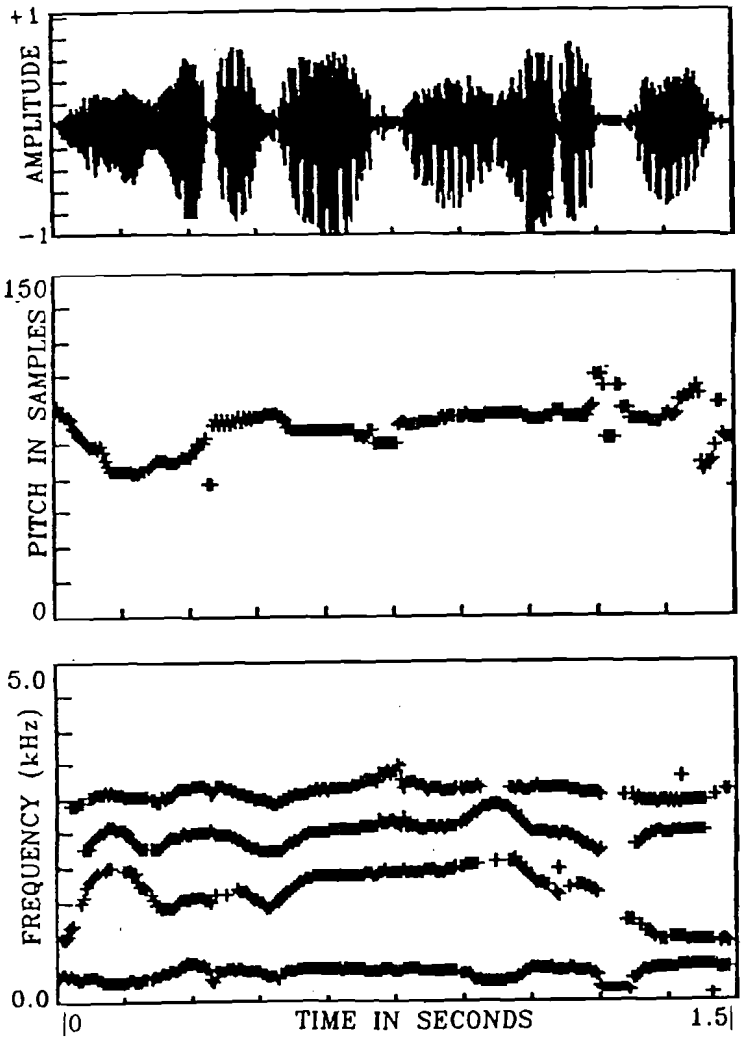
The additive noise in equation (4.19) introduces new zeros and redistributes the zeros of the given signal. If the noise is not too high the modified group delay function can be used to estimate the pitch period and location of formants from noisy speech also. Fig.4.17 shows the utterance of Fig.4.16 corrupted by white Gaussian noise and the corresponding pitch and formant data. The overall SNR is 3dB. The SNR as a function of time is superimposed on the pitch and formant data.

4.5.2 Speech Synthesis

The formant and pitch data obtained in the previous section are used in the formant vocoder (discussed in Chapter 3) to synthesise speech. The formant bandwidths are fixed as a percentage of the formant. Although there is a significant difference in intelligibility it is observed that the naturalness is almost completely lost.

4.6 Summary

In this Chapter we have proposed a new technique for processing the Fourier transform phase spectrum of the speech signal to estimate the parameters corresponding to the vocal tract system and excitation source. The standard phase spectrum is considered difficult to

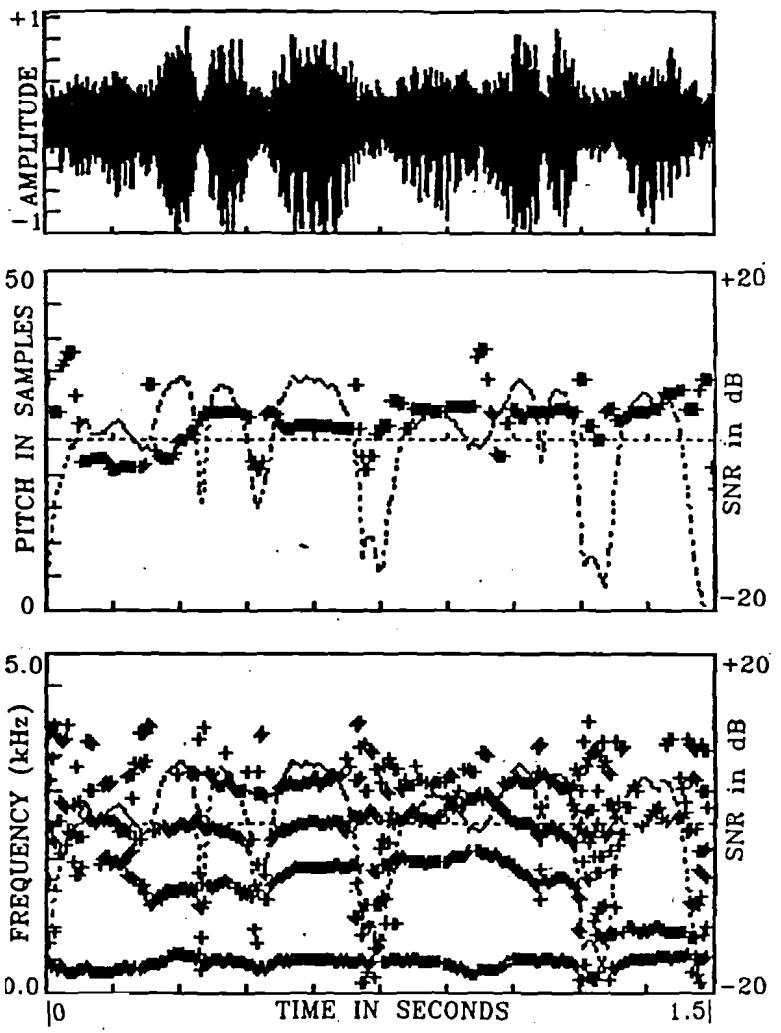


(a) Signal

(b) Pitch data

(c) Formant Data

Fig. 4.16 Formant and pitch extraction from natural speech.



(a) Signal

(b) Pitch data

(c) Formant Data

Fig. 4.17 Formant and pitch extraction from noisy speech.

interpret due to the artifacts introduced by the zeros of the z-transform of the excitation function and data windows. We have proposed a technique to process the phase in which the effect of zeros is significantly reduced. The main results of this study are:

(1) The fluctuations caused by zeros are reduced.

(2) The effects of time window functions are significantly reduced.

(3) The most significant result is that it seems possible to estimate both formants and pitch from natural speech even at low SNRs.

(4) Although it is not possible to estimate the bandwidths corresponding to formants it is still possible to synthesise intelligible speech from formant and pitch data alone.

CHAPTER 5

SPECTRUM ESTIMATION USING MODIFIED CROUP DELAY FUNCTIONS

5.1 Introduction

Estimation of the power spectral density (PSD) or simply spectrum of discretely sampled deterministic and stochastic processes is usually based on procedures employing the Discrete Fourier Transform (DFT). Specifically the PSD of a discrete time signal may be defined as follows :

Consider a discrete time deterministic signal x(n) which is complex valued and absolutely summable, i.e. the signal energy is finite, then

$$\mathcal{E} = \sum_{n=-\infty}^{\infty} |x(n)|^2 \leq \infty \qquad (5.1)$$

the DFT exists and is defined by

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n) e^{-jwn} \qquad (5.2)$$

The square modulus of the DTFT $X(\omega)$ is often termed the spectrum of x(n) or

$$S(\omega) = |X(\omega)|^2 \qquad (5.3)$$

Parseval's theorem

$$\sum_{n=-\infty}^{\infty} |x(n)|^2 = \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \qquad (5.4)$$

is a statement of the principle of conservation of energy; the energy of the time signal is equal to the energy of the frequency domain transform $\int_{-\infty}^{\infty} S(\omega) d\omega$. Thus $S(\omega)$ is an energy spectral density in that it represents the distribution of energy as a function of frequency.

For the case when x(n) is a realisation of a stationary random process, the power spectrum is computed indirectly through the autocorrelation function. The Wiener Khinchin theorem states the relation between power spectrum $S(\omega)$ and the autocorrelation

sequence $R_{xx}(m)$ corresponding to the discrete time sequence $x(n)$:

$$R_{xx}(m) = E[x(n+m)x^*(n)] \tag{5.5}$$

$$S(\omega) = \sum_{m=-\infty}^{w} R_{xx}(m)e^{-jom} \tag{5.6}$$

where E is the expectation operator.

In practice the statistical autocorrelation function is not known. An additional assumption that is made is that, the process is ergodic in the first and the second moments and equation (5.5) is replaced by a time average rather than ensemble average :

$$R_{xx}(m) = \sum_{n=-\infty}^{w} x(n+m)x^*(n) \tag{5.7}$$

For both deterministic and non deterministic signals, the data is available only at a finite number of points, namely, $x_0, \ldots, x_{N-1}$ and the estimate of $S(\omega)$ is given by

$$S(\omega) = \frac{1}{N}|X(\omega)|^2 \tag{5.8}$$

or

$$S(\omega) = \frac{1}{N} \sum_{m=0}^{N-1} R_{xx}(m)e^{-jom} \tag{5.9}$$

and $S(\omega)$ is defined in the interval $[-\pi, +\pi]$. In practice $R_{xx}(m)$ is replaced by (5.7) for nondeterministic signals and $S(\omega)$ reduces to equation (5.8).

Thus the estimation of the power spectral density reduces to one of esimation of the FT magnitude spectrum. When the discrete Fourier transform (DFT) is used to compute the PSD, the estimated PSD $\hat{S}(\omega_k)$ is available only at a discrete set of frequencies i.e.

$$\hat{S}(\omega_k) = S(\omega)|_{\omega=2\pi k/N} \tag{5.10}$$

This spectrum estimate is called the **periodogram.**

The finite length requirement for practical signals means that the signal is multiplied by a rectangular window, and the overall transform is a convolution of the true transform with that of the window transform. If the true power spectrum is concentrated in a

narrow bandwidth, this convolution will spread the spectrum into adjacent frequency regions. This phenomena is called leakage. Thus the data window is a primary factor that determines the frequency resolution of the periodogram. Leakage effects are reduced by an appropriate choice of windows with non uniform weighting. But the price paid for a reduction in the sidelobes is always a broadening in the main lobe of the window transform, which in turn decreases the resolution of the spectral estimate.

In the parametric approach to spectrum estimation a model is used to extrapolate the data outside the window. It is then usually possible to obtain a better spectral estimate based on the model by determining the parameter of the model from the observation.

The spectrum analysis in the context of modelling becomes a three step procedure : (i) selection of a model, (ii) estimation of parameters for the assumed model and (iii) obtain spectral estimates from the model parameters. The degree of improvement in resolution and spectral fidelity, if any will be determined by the ability to fit an assumed model with a few parameters derived from the measured data.

Although there are a number of performance advantages that may be obtained using model based methods, the advantages strongly depend upon the signal-to-noise ratio (SNR) as might be expected. In fact for low SNRs, the spectral estimates based on modelling are no better than those obtained using conventional DFT processing.

In this Chapter we suggest an alternative to the periodogram approach to spectrum estimation. This method is based on the modified group delay function defined in Chapter 4. In Chapter 4 we saw that the modified group delay function could be successfully used to estimate parameters from natural signals like speech. We also saw in Chapter 4 that the modified group delay function was also

successful in estimating parameters from noisy speech. It is worth noting that no assumptions (about the mechanism for generating the signal) were explicitly used in the algorithm for estimating the modified group delay function. In this Chapter we show that the modified group delay functions can be used to address the general problem of spectrum estimation.

In Section 5.2 we discuss the proposed method of spectrum estimation using modified group delay functions. In Section 5.3 we demonstrate the results of this approach to spectrum estimation through several illustrative examples. In particular we consider two examples, namely, (i) estimation of sinusoids in noise and (ii) estimation of narrow band auto-regressive (AR) processes in noise. In Section 5.4 we compare the performance of the proposed method of spectrum estimation with that of the periogram approach. Resolution is primarily dictated by the size of the data window as per the standard time bandwidth product relation.

5.2 Principle of the Method:

As mentioned before, our objective is to estimate the spectral features of an autoregressive process or a sinusoidal process in noise using the properties of Fourier transform phase, or equivalently using group delay functions.

Let us consider the output x(n) of an autoregressive process s(n) corrupted with noise u(n).

That is

$$x(n) = s(n) + u(n) \tag{5.11}$$

where S(z) the z-transfrom of s(n) is obtained as

$$S(z) = \frac{GE(z)}{A(z)} \tag{5.12}$$

E(z) is the z-transform of the excitation sequence e(n), where e(n) is white Gaussian noise with variance unity, and G/A(z) is the

all-pole system corresponding to the autoregressive process. Now

$$X(z) = \frac{GE(z) + U(z)A(z)}{A(z)} = \frac{V(z)}{A(z)} \qquad (5.13)$$

The group delay function of X(z) in terms of the group delay functions of V(z) and A(z) is given by

$$\tau_X(\omega) = \tau_V(\omega) - \tau_A(\omega) \qquad (5.14)$$

The Fourier transform of x(n) is given by

$$X(\omega) = \frac{GE(\omega) + A(\omega)U(\omega)}{A(\omega)} \qquad (5.15)$$

For low noise levels the first term $GE(\omega)$ dominates and hence the group delay function $\tau_X(\omega)$ of X(z) behaves almost like the group delay function $\tau_S(\omega)$ of S(z) (noise free case). For high noise levels two cases have to be considered separately: (a) Regions (say $\bar{R}$) of frequency where the values of $|A(\omega)|$ are not small (i.e, not near zero) and also the shape of $|A(\omega)|$ curve is smooth, and (b) Regions (say R) of frequencies where the values of $|A(\omega)|$ are so small that the first term in V(z), namely GE(z), dominates. In regions $\bar{R}$ the group delay function $\tau_V(\omega)$ corresponding to the numerator polynomial of Eq (5.13) behaves like for any noise sequence. That is, there will be large positive and negative spikes depending on the roots of V(z) in the region $\bar{R}$. In the regions R the group delay function $\tau_V(\omega)$ still will have large amplitude spikes of either polarity, but this time they are contributed by the roots of V(z) in the region R, where the first term in V(z) dominates. Thus in both the regions $\bar{R}$ and R the group delay function behaves like that for a noise sequence, but due to different sources of noise. The most important point is that the spiky nature of the group delay function $\tau_X(\omega)$ is not affected significantly by the presence of A(z) in the numerator. This is the reason why the first term $\tau_V(\omega)$ in $\tau_X(\omega)$ is distinct from the second term $\tau_A(\omega)$. So the characteristics

of the second term can still be estimated by suppressing the spikes in the overall group delay function $\tau_x(\omega)$. That this works even for very low noise levels is obvious from this argument.

The basis for our new spectrum estimation procedure is to suppress the large amplitude spikes in $\tau_x(\omega)$ due to $\tau_v(\omega)$ in order to highlight the desired components $\tau_A(\omega)$. To suppress the spikes due to noise, it is necessary to identify their locations and then reduce their amplitudes. To do this we can take advantage of the modified group delay function derived in Chapter 4. The modified group delay function in the context of spectrum estimation is used to suppress the zeros that are introduced by additive noise and the zeros that are introduced by the data window. If the modified group delay function can be thought of as an approximate estimate of the group delay function corresponding to that of a minimum-phase system, the relationship between group delay functions for minimum phase signals can be used to estimate the spectrum of the given signal.

For minimum phase signals we saw in Chapter 2 that $\tau_p(\omega) = \tau_m(\omega)$, where $\tau_p(\omega)$ and $\tau_m(\omega)$ are the group delay functions derived from the phase and magnitude respectively. Using the relationship between the cepstral coefficients and the group delay function, the spectrum can be derived. Table.S.l gives the algorithm for computing the spectrum using modified group delay functions.

5.3 Illustrations

We consider two types of problems for illustration.

Example-:: Autoregressive process in noise (estimation of the AR spectrum)

$$x_1(n) = s(n) + u(n) \qquad (5.16)$$

$$s(n) = -\sum_{k=1}^{4} a_k s(n-k) + Ge(n) \qquad (5.17)$$

where the excitation e(n) is white Gaussian noise of variance unity

Table 5.1 Algorithm for computing the spectrum from the Modified group delay function $\tau_o(k)$.

1. Compute the estimate of the weighted cepstrum from $\tau_o(k)$ as follows. Compute N-pt IDFT of $\tau(k)$

$$\hat{c}(n) = IDFT[\tau(k)], n = 0,1,...,N-1.$$

2. Form the sequence $c_1(n)$

$$\left.\begin{aligned} c_1(0) &= 0 \\ c_1(n) &= \hat{c}(n)/n \\ c_1(N-n+1) &= c_1(n) \end{aligned}\right\}, 1 \le n \le N/2.$$

3. Compute the N-pt DFT of $c_1(n)$

$$X_1(k) = DFT[c_1(n)], \quad k = 0,1,...,N-1$$

4. Compute

$$\ln|X_s(k)| = Real[X_1(k)].$$

$2*\ln|X_s(k)|$ is the estimated smoothed spectrum as obtained from the modified group delay.

and $u(n)$ is an additive noise with variance dependent upon the desired signal-to-noise ratio (SNR). The values of the coefficients are: $a_1 = -2.760$, $a_2 = 3.809$, $a_3 = -2.654$ and $a_4 = 0.924$.

Example-2: Two sinusoids in noise (estimation of frequencies of the sinusoids)

$$x_2(n) = \sqrt{10}\exp[j2\pi(0.10)n] + \sqrt{20}\exp[j2\pi(0.15)n] + u(n) \qquad (5.18)$$

where $u(n)$ is an additive white Gaussian noise with the variance dependent upon the SNR. These examples are similar to the ones used in [S.M.Kay; 1988] for discussion of periodogram estimates. We assume a sampling frequency of 10kHz and number of samples N=256 for Example-1, and N=100 for Example-2. Different realizations of $x_1(n)$ and $x_2(n)$ are obtained by using a different noise sequence each time.

Figs.5.1, 5.2 and 5.3 give the periodogram, group delay function and the new magnitude spectrum estimates of the autoregressive process from the noisy signal (SNR = 20dB) of Example-1. Figs. 5.1a,

5.2a and 5.3a show the plots for a single realization of clean the data. Figs. 5.lb, 5.2b and 5.3b show the plots for 50 realizations of the data. Figs. 5.lc, 5.2c and 5.3c show the averaged plots. It is to be noted that, as expected, periodogram estimate has large variance (Fig.5.1b). Reduction of fluctuations by averaging several periodograms introduces large bias [S.M.Kay; 1988]. 'The fluctuation is significantly reduced in the estimated group delay'functions and the spectra estimated from the group delay function (Figs.5.2b, 5.3b).Figs. 5.la, 5.2a and 5.3a show that it is possible to reduce the fluctuations even by processing a single realization. In fact a single realization seems to restore all the information that can be obtained from averaged plots (Compare Fig. 5.3a with 5.3c). Note also that averaging reduces the dynamic range in periodogram (Figs. 5.1a and 5.1c) whereas averaging group delay functions does not seem to affect the dynamic range (Figs. 5.3a and 5.3c).

Although we have not discussed the theory, we have applied our method for estimating sinusoids in noise. The results are shown in the plots given in Figs. 5.4, 5.5 and 5.6 for SNR = 20dB. Our method works well even for estimating sinusoids in the presence of noise. The same general conclusions as for the autoregressive process hold good for sinusoidal process regarding variance and bias of the estimates.

Note that the finite data window also produces large spikes in the group delay function. But division by the cepstrally derived smoothed magnitude spectrum suppresses the sidelobe effects of the window also. This way the estimated magnitude spectrum from the group delay function is less dependent on the window. However the resolution of the spectral peaks is dependent on the size of the window and that effect can be seen in the estimated spectrum from the group delay function. Fig.5.7shows the plots for noise free

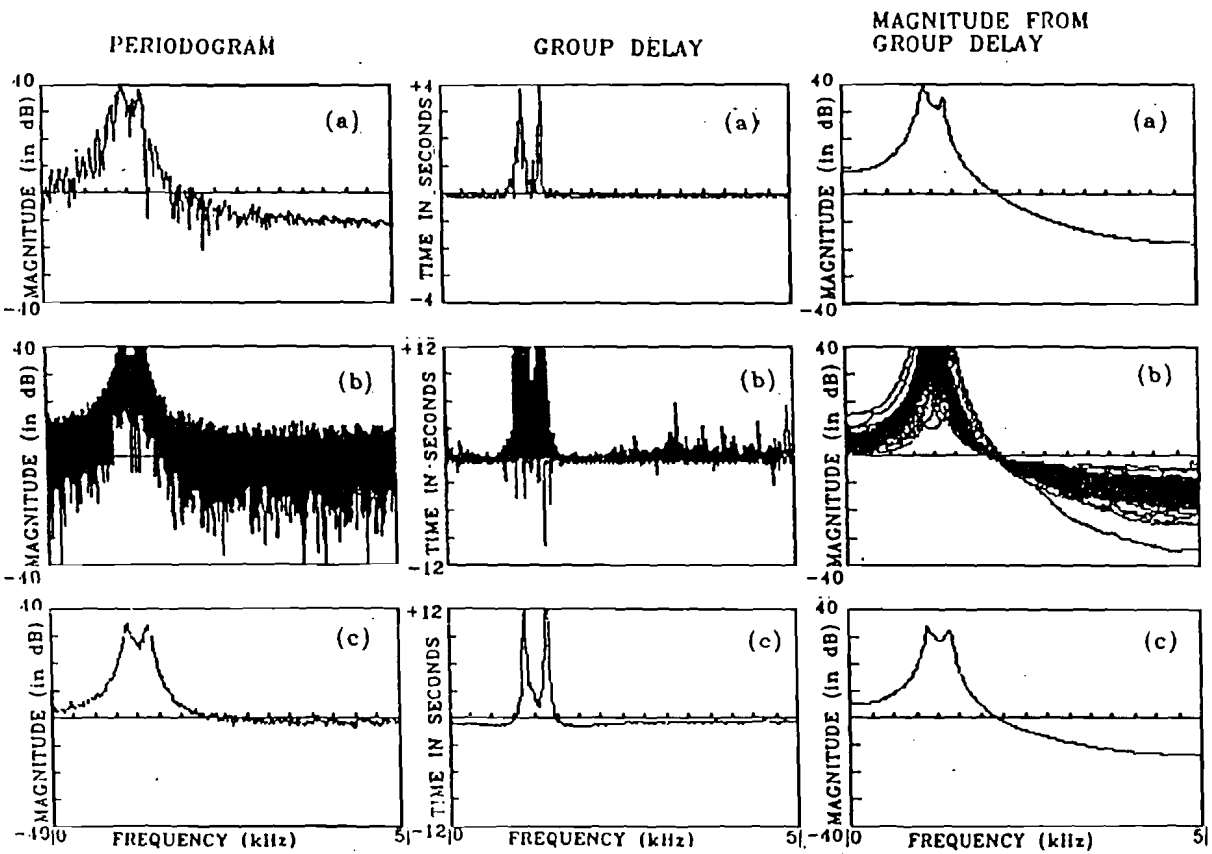


Fig.5.1 Periodogram estimate of spectrum for an AR process in noise (a) single realisation (clean data) (b) 50 overlaid realisations (SNR = 20 dB) and (c) Average of realisations.

Fig.5.2 Estimated group delay function for an AR process in noise (a) single realisation (clean data) (b) 50 overlaid realisations and (c) Average of realisations

Fig.5.3 Estimated spectrum from group delay function for an AR process in noise (a) single realisation (clean data) (b) 50 overlaid realisations and (c) Average of realisations.

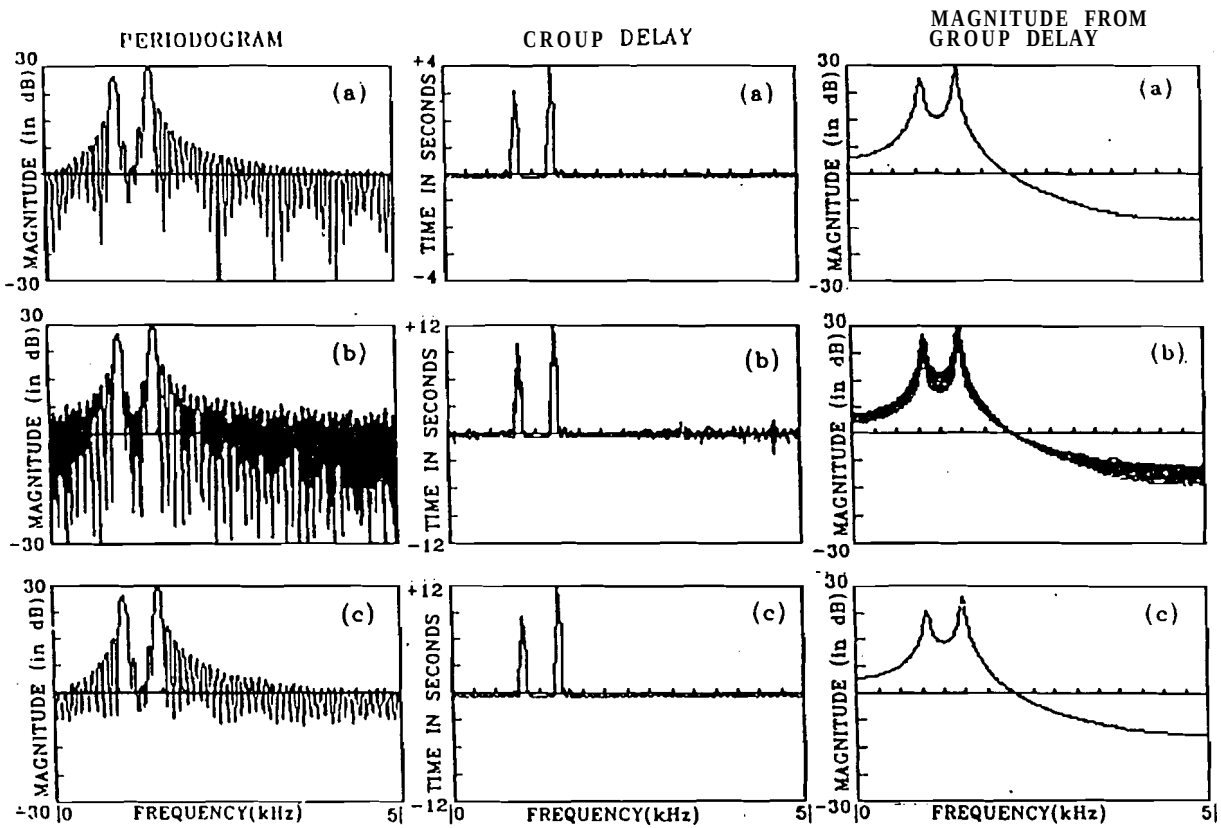


Fig.5.4 Periodogram estimate of spectrum for sinusoids in noise (a) single realisation (clear data) (b) 50 overlaid realisations (SNR = 20 dB) and (c) Average of realisations.

Fig.5.5 Estimated group delay function for sinusoids in noise (a) single realisation (cleannoise data) (b) 50 overlaid realisations and (c) Average of realisations.

Fig.5.6 Estimated spectrum from group delay function for sinusoids in noise (a) single realisation (clean data) (b) 50 overlaid realisations and (c) Average of realisations.

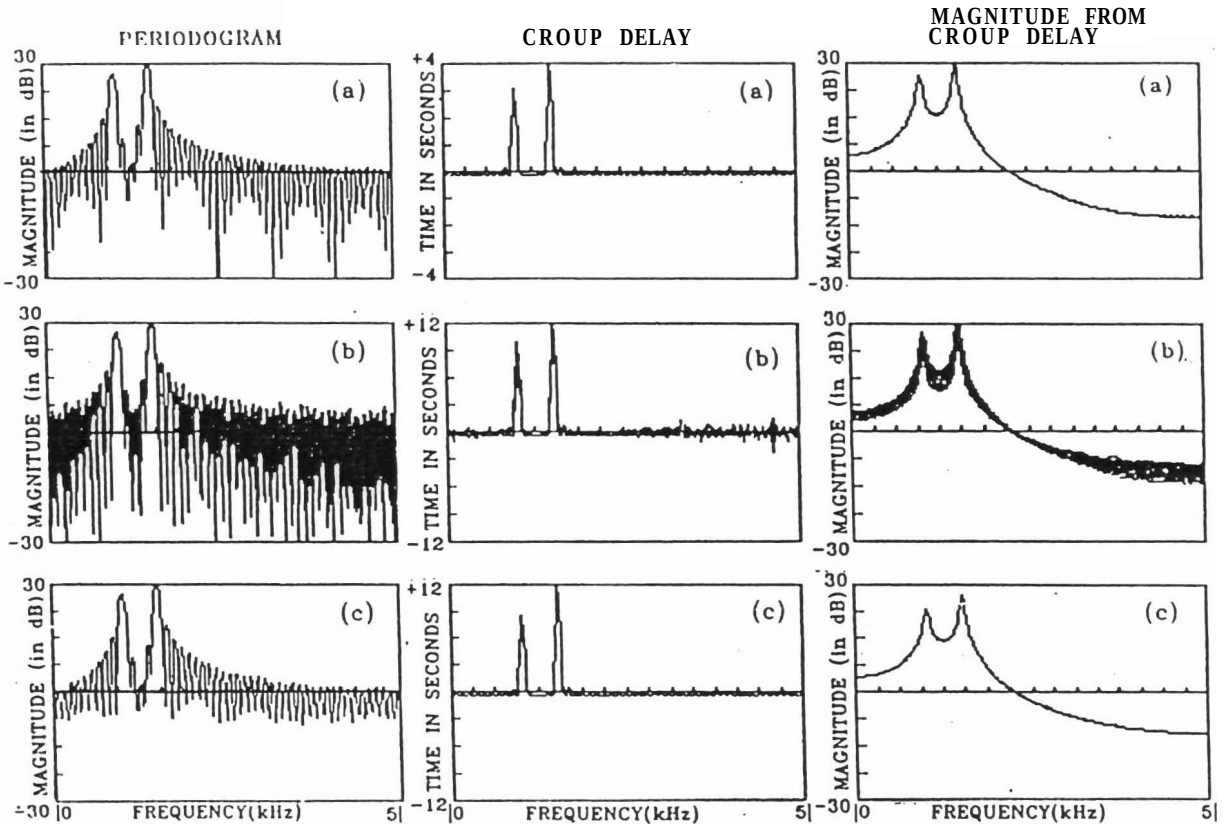


Fig.5.4 Periodogram estimate of spectrum for sinusoids in noise (a) single realisation (clear data) (b) 50 overlaid realisations (SNR = 20 dB) and (c) Average of realisations.

Fig.5.5 Estimated group delay function for sinusoids in noise (a) single realisation (clean data) (b) 50 overlaid realisations and (c) Average of realisations.

Fig.5.6 Estimated spectrum from group delay function for sinusoids in noise (a) single realisation (clean data) (b) 50 overlaid realisations and (c) Average of realisations.

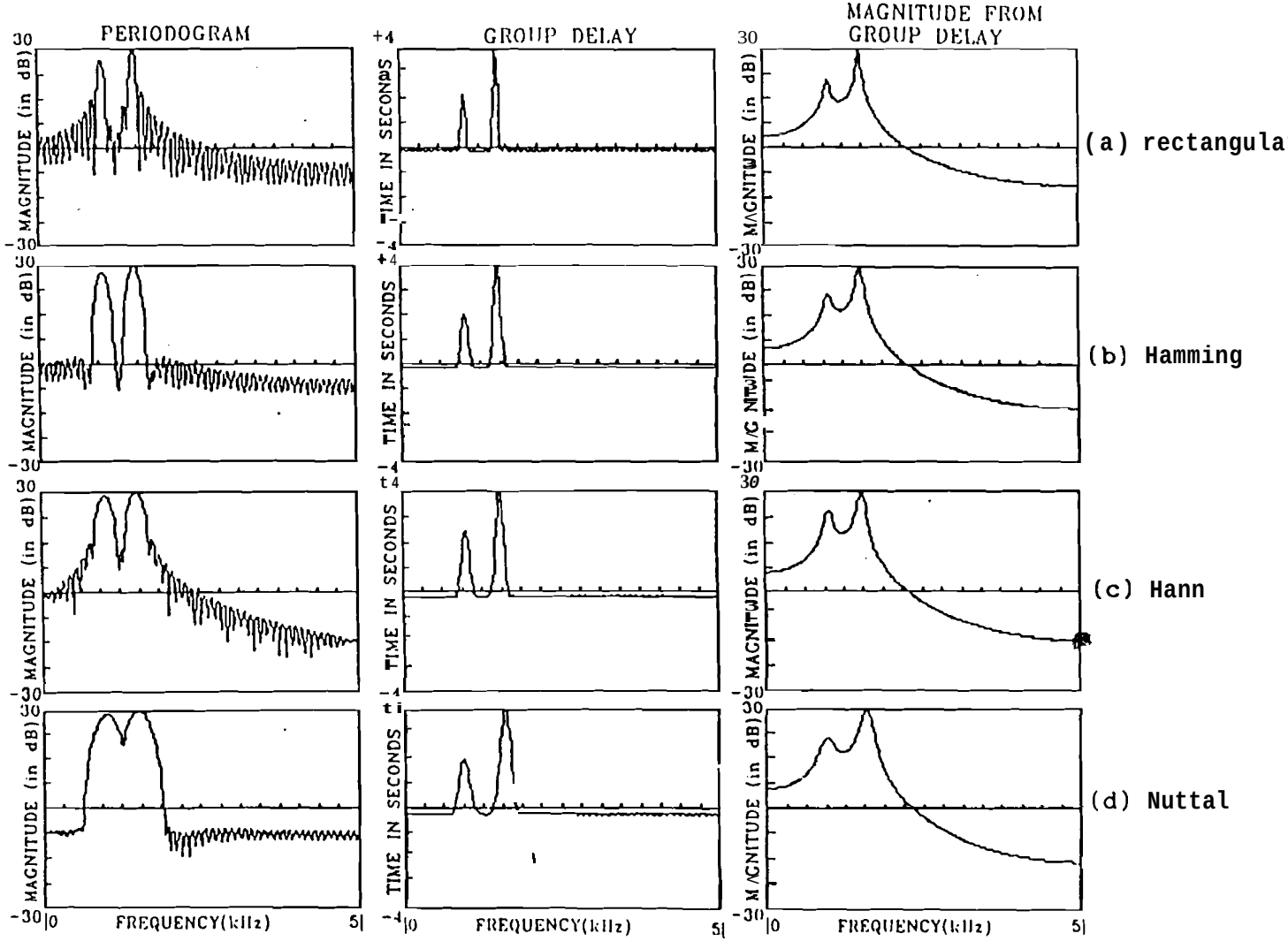


Fig.5.7 Effect of type of data window on the estimated, spectra.

sinusoidal data for different types of windows (Rectangular, Hamming, Hann and Nuttall) [S.L.Marple;1987]. While the sidelobe effects are reflected in the periodogram spectrum (Fig.5.7a) in both the dynamic range as well as in the width of the mainlobe, the corresponding group delay function plots (Fig.5.7b) do not seem to be affected by the sidelobes. THe effective window size is reflected in the width of the spikes in the group delay function, with smallest width for the rectangular window and largest width for Nuttall window. The window size effect can be more explicitly seen in Fig.5.8, where the plots are given for different sizes (512, 128, 32, 16 and 8 respectively) of the rectangular window. As before the window size seems to affect the width of the peaks in the group delay functions But the sidelobe effects are almost suppressed.

What is achieved by the new method is that we can estimate a spectrum with fluctuations suppressed, preserving the resolution properties of the periodogram estimate. The frequency resolution limit is set by the data window size. We can see the effect of the window size on the frequency resolution in Fig.5.9, where the periodogram, group delay and derived magnitude spectrum are shown for different spacings of frequencies of two sinusoids of equal amplitudes using 128 samples of the data. Note that upto 60 Hz separation, the two frequencies are resolved with 12.8ms of data.

Fig.5.10 shows the plots for the sinusoids with different amplitudes. The fluctuations due to sidelobes are reduced even when the amplitudes are significantly different. That is, the periodogram resolution features are reflected in the group delay function without sidelobe effects.

Figs.5.11 and 5.12 show the results of the estimated spectra for different noise levels (SNR = 10dB, SNR = 0dB, SNR = -10dB). In Figs.5.11 and 5.12 the plots for noisy data are presented as an

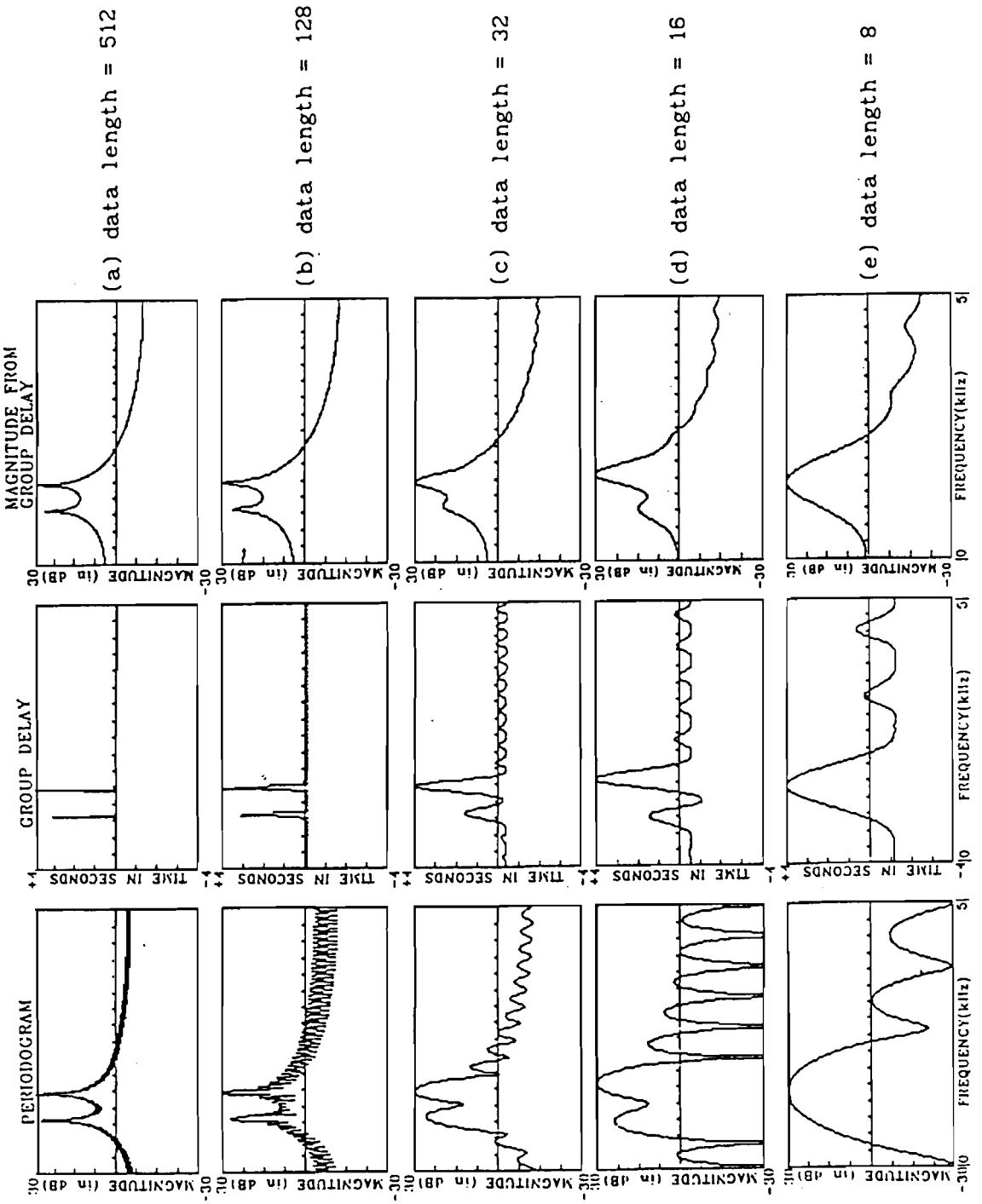


Fig.5.8 Effect of size of rectangular window on the estimated spectra

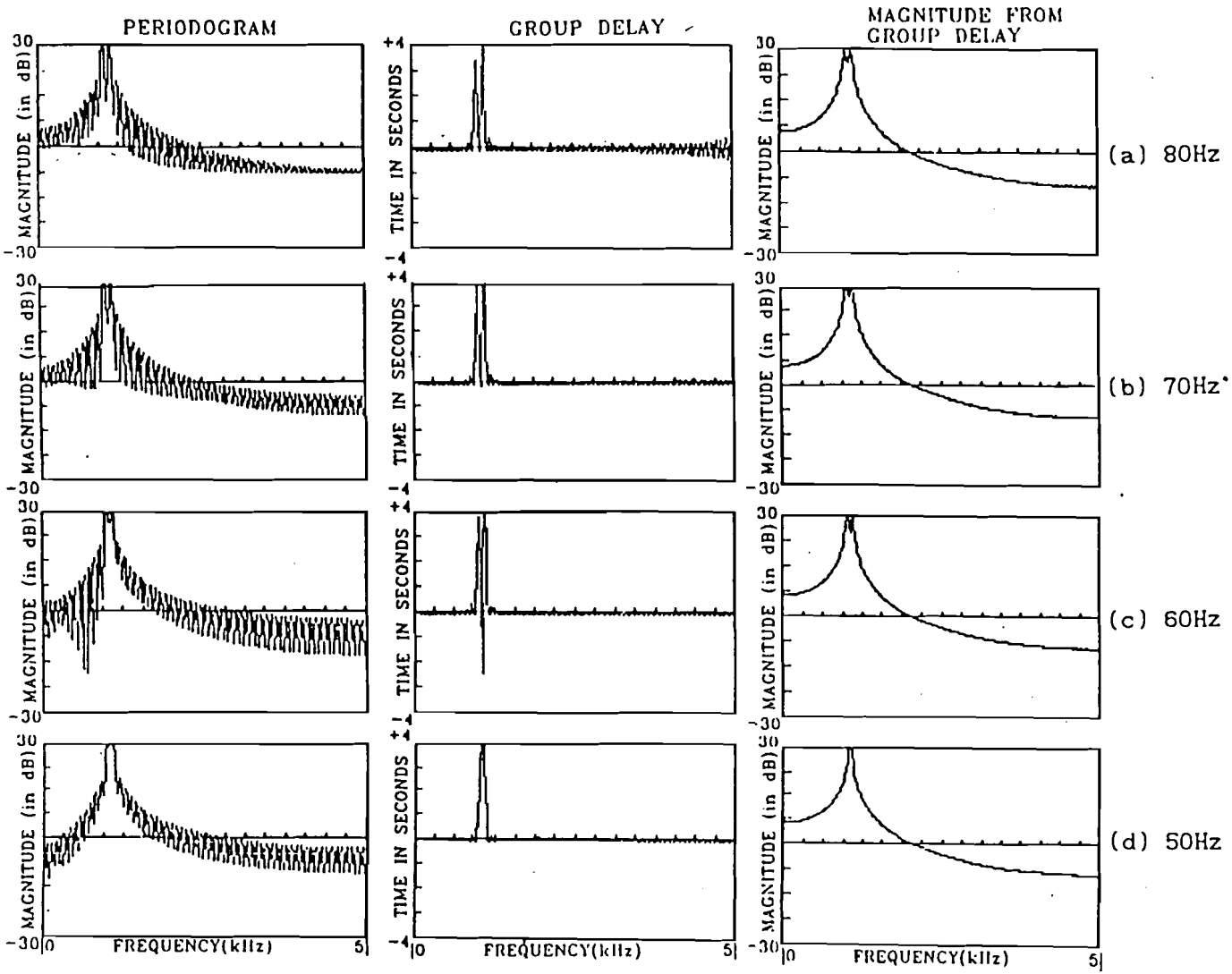


Fig.5.9 Effect of rectangular window size on frequency resolution in the estimated spectra. Data consists of 128 samples of two sinusoids.

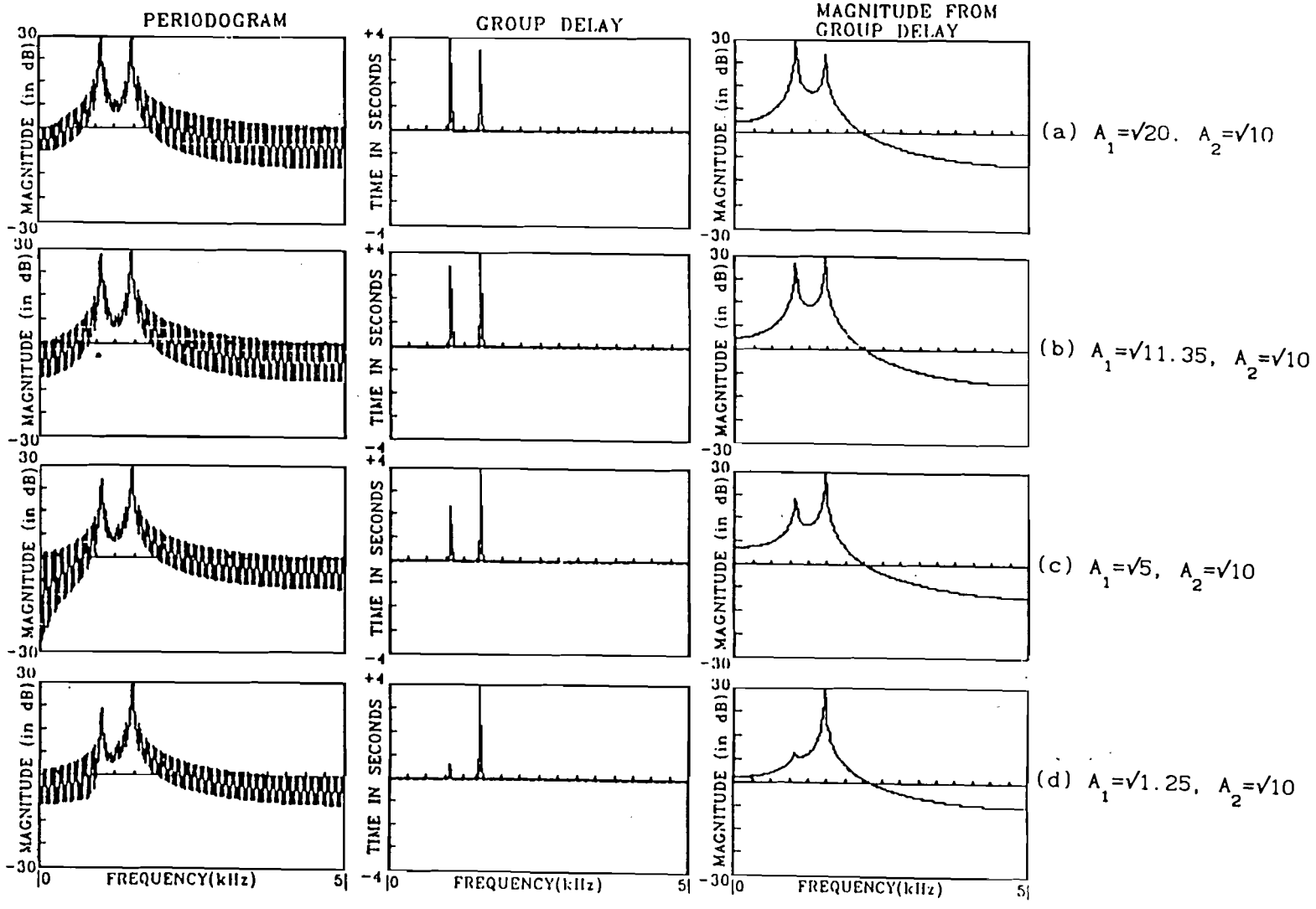
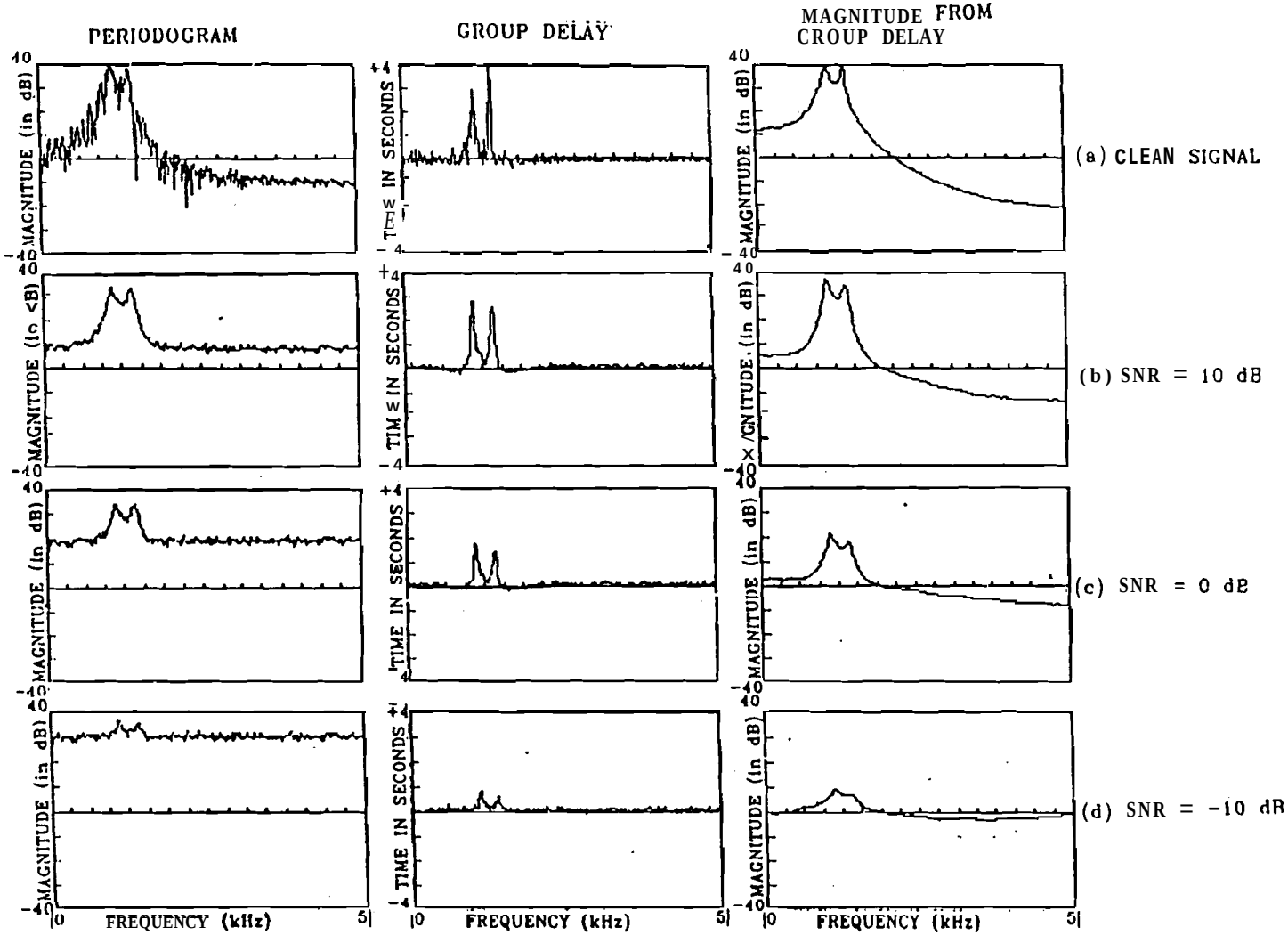


Fig. 5.10 Estimated spectra for different amplitudes of sinusoids. Data consists of 256 samples of two sinusoids with amplitudes A_1 and A_2 separated by 500Hz.



Flg.5.11 Estimated spectrum from group delay functions for dlfferent noise levels for an AR process in noise. For noisy data the spectra are presented as an average over 50 reallsations.

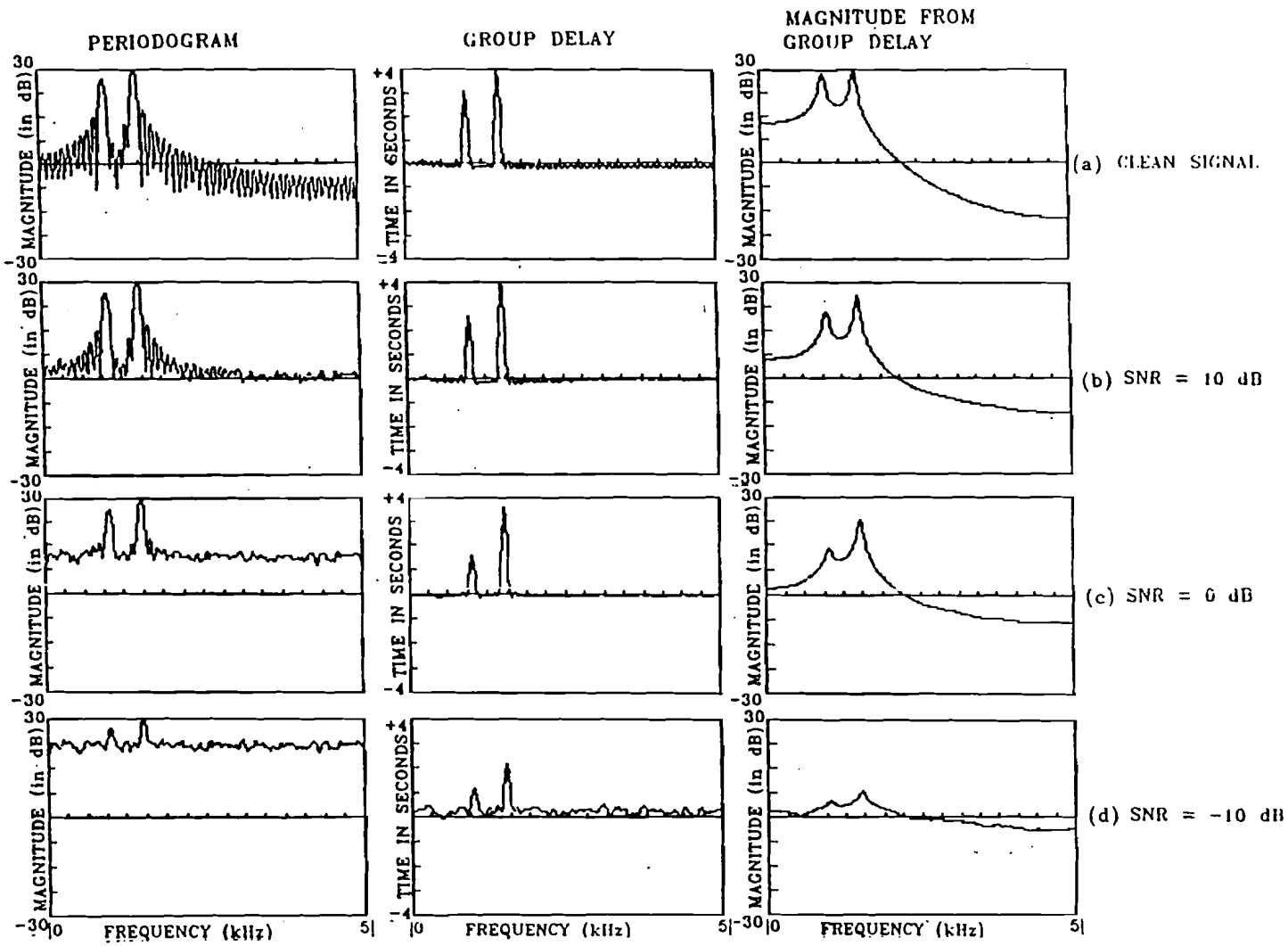


Fig.5.12 Estimated spectrum from group delay functions for different noise levels for sinusoids in noise. For noisy data the spectra are presented as an average over 50 realisations.

average of 50 realisations. The features are restored with few spurious peaks in the averaged spectral plots shown in Figs 5.11 and 5.12. The desired spectral features are seen even for SNR = -10dB. These results show that the proposed method works even for high noise levels. Note that model based AR spectrum estimation will not work for noisy data [S. M. Kay; 19881. Fig.5.13 gives a comparison of the performance of our method of spectrum estimation with Burg's method[S. M. Kay; 19881. The data consists of 256 samples of AR process in noise. The Burg's method uses an 8th order model. Note that the group delay function preserves the resolution properties of the periodogram, with much less fluctuations, even for low SNR. Unlike periodogram spectrum, the group delay method restores the dynamic range of the AR spectrum even at high noise levels. Model based techniques fail to resolve the peaks at high noise levels (SNR < 5dB). If the order of the model is increased, more spurious peaks will be generated. Superiority of our method in resolving peaks and reducing spurious peaks is evident from the figure even for low values of SNR.

5.4 *Bias-Variance Calculations*

It is difficult to obtain analytical expressions for bias and variance for the spectrum estimated using modified group delay functions. In the spectrum estimated using the modified group delay functions the scale factor is lost as the value of c(0), the zeroth cepstral coefficient is not available.

To get a feel for the bias the averaged periodogram estimates and group delay spectrum estimates are obtained as follows. If $S_N(\omega)$ is the estimated spectrum using a datalength of N, the average of 50 realisation is obtained as

$$\bar{S}(\omega) = \frac{1}{N_R} \sum_{i=1}^{N} S_N^i(\omega) \qquad (5.19)$$

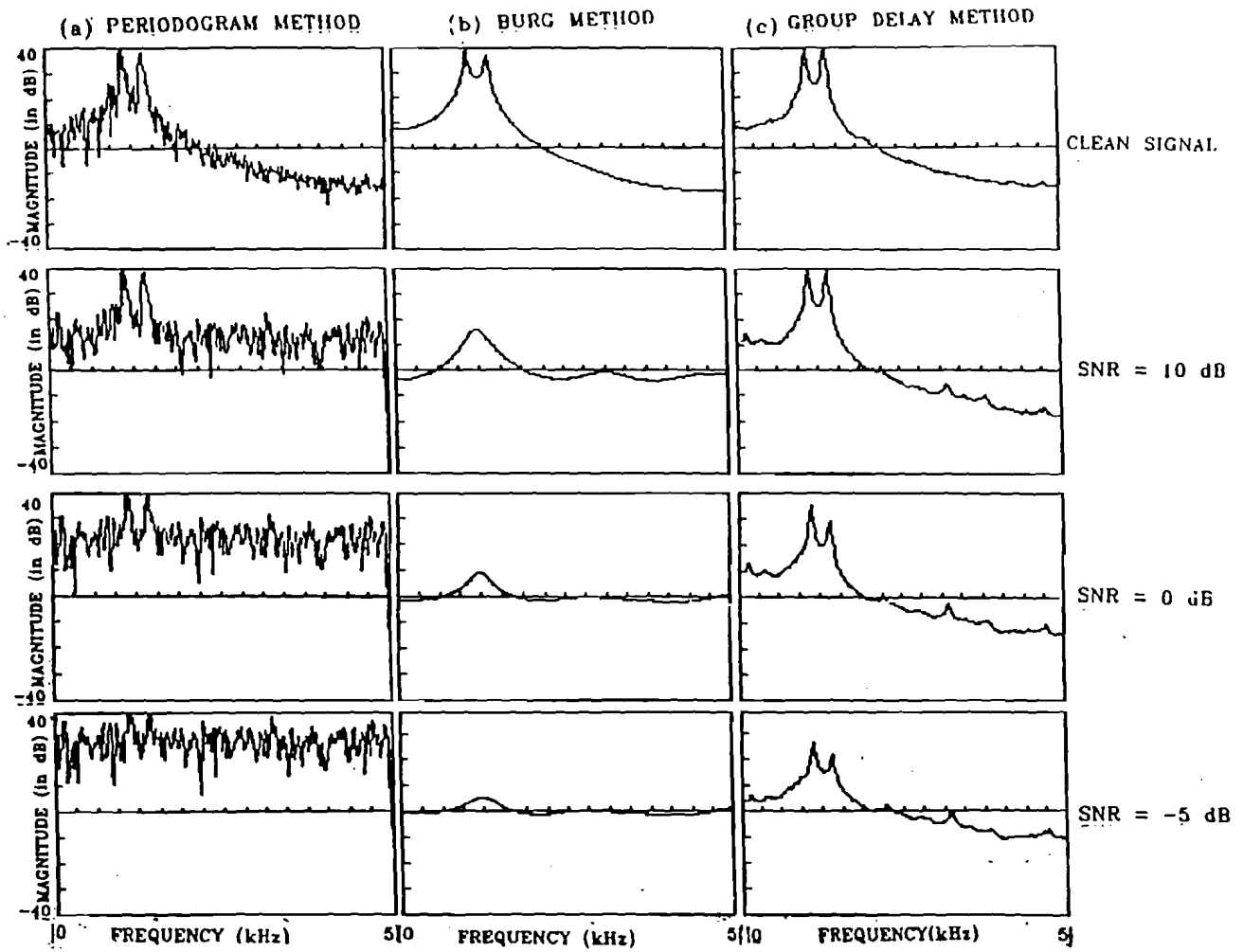


Fig.5.13 Comparison of the group delay spectrum estimation technique with model-based Burg method for different noise levels for an AR process in noise.

where is $S_N^i(\omega)$ is the estimated spectrum for a realisation and N_R is the number of realisations. In the examples given, N_R is 50. In the examples that follow we only consider the AR process in noise. This is because it is possible to compute the true **AR** spectrum from the known **AR** coefficients. We superimpose the true **AR** spectrum plot on each of the estimated spectra. We consider two different cases for illustration (i) Comparison of the proposed method with' that of the periodogram approach for different datalengths for AR process in noise and (ii) Comparison of the proposed method with that of the periodogram approach for different noise levels for a fixed datalength (256 samples). It is observed that there is hardly any significant difference between the group delay method and the periodogram method for different datalengths as indicated in Fig.5.14. For different noise levels we observe that the bias in the estimates for the group delay method are much less than that of the estimates obtained using the periodogram approach. This is illustrated in Fig.5.15.

For variance calculations, the variance of estimate was computed as :

$$\sigma^2(\omega) = \frac{1}{N_R} \sum_{i=1}^{N} \left[S_N^i(\omega) - \bar{S}(\omega) \right]^2 \qquad (5.20)$$

where $\bar{S}(\omega)$ is the average of the spectral estimate obtained through Eq(5.19). We now superimpose the plots of the estimated variance for the periodogram and group delay spectrum for (i) different datalengths of **AR** process (noiseless case) (ii) AR process in noise for a fixed datalength (256 samples) and (iii) sinusoids in noise for a fixed datalength (256 samples). For both the cases of the **AR** process we find that the variance of the group delay derived spectrum estimates is considerably larger than that of the variance of the periodogram estimates especially in the region corresponding to the

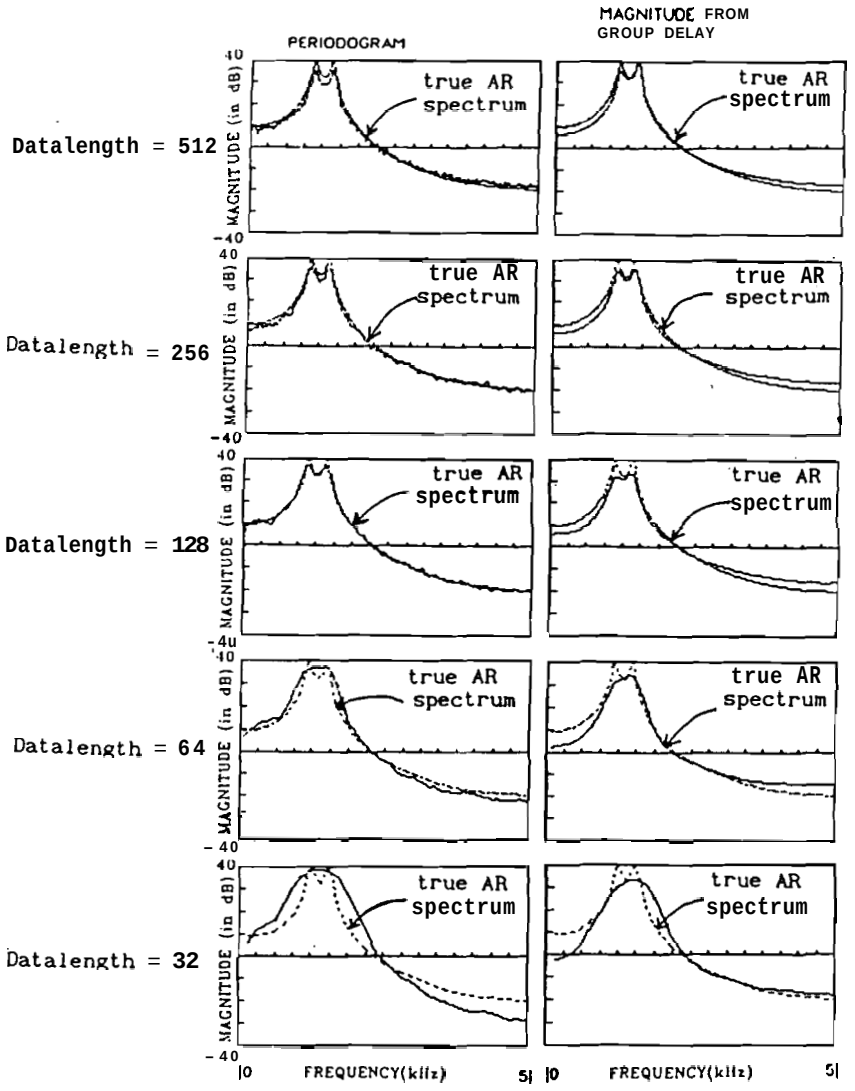


Fig.5.14. Illustration of the bias of estimates for AR process (clean) for different data lengths. The true AR spectrum is superimposed on all the plots. The results for the periodogram and group delay spectra are presented as an average of 50 realisations.

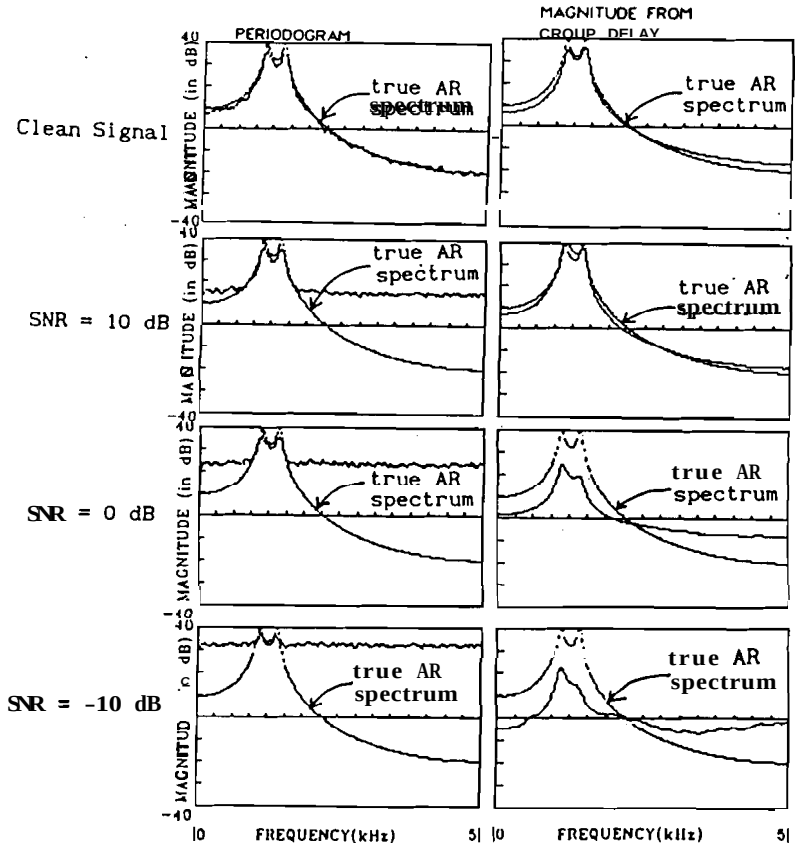


Fig.5.15. Illustration of the bias of estimates for AR process in noise for different noise levels. The datalength is 256 samples. The true AR spectrum is superimposed on all the plots. The results for the periodogram and group delay spectra are presented as an average of 50 realisations.

location of resonances (Figs.5.16-5.17). In fact, the behaviour is not consistent in that it seems to be neither dependent on datalength or SNR. For sinusoids in noise the variance is significantly lower for the MGD estimate than that of the peridogram estimate for low noise levels. For high noise levels, for example -10 dB we observed that the variance of the estimated spectrum using group delay method become larger than that of the periodogram estimates for the same noise level as shown in Fig.5.18.

The difference in the behaviour of the variance for the case sinusoids and AR process may be due to the location of the zeros. For sinusoids the window zeros are uniformly distributed around the unit circle in the z-domain. As long as the window zeros are not significantly disturbed, the variance is low. At high'noise levels, for examples, -10 dB, there is a possibility that the window zeros are significantly disturbed. This results in large variance of the spectrum derived from group delay as indicated in Fig.5.18. For the AR process, the excitation is Gaussian noise and the zeros are not uniformly distributed in the z-domain. As mentioned in Chapter 4 the proposed method works provided that a zero (which may be due to the excitation, noise and window effects) does not lie close to a resonance. For .thecase of sinusoids in noise it is possible to choose an appropriate window apriori (both shape and size) as the location of window zeros in the z-domain does not change for different realisations. For the AR process as the excitation is Gaussian noise it difficult to choose the window shape and size apriori, as the distribution of excitation is quite random in z-domain.

5.5 Summary

In summary we have proposed a method of spectrum estimation that (a) reduces fluctuations caused by the variance of noise and window

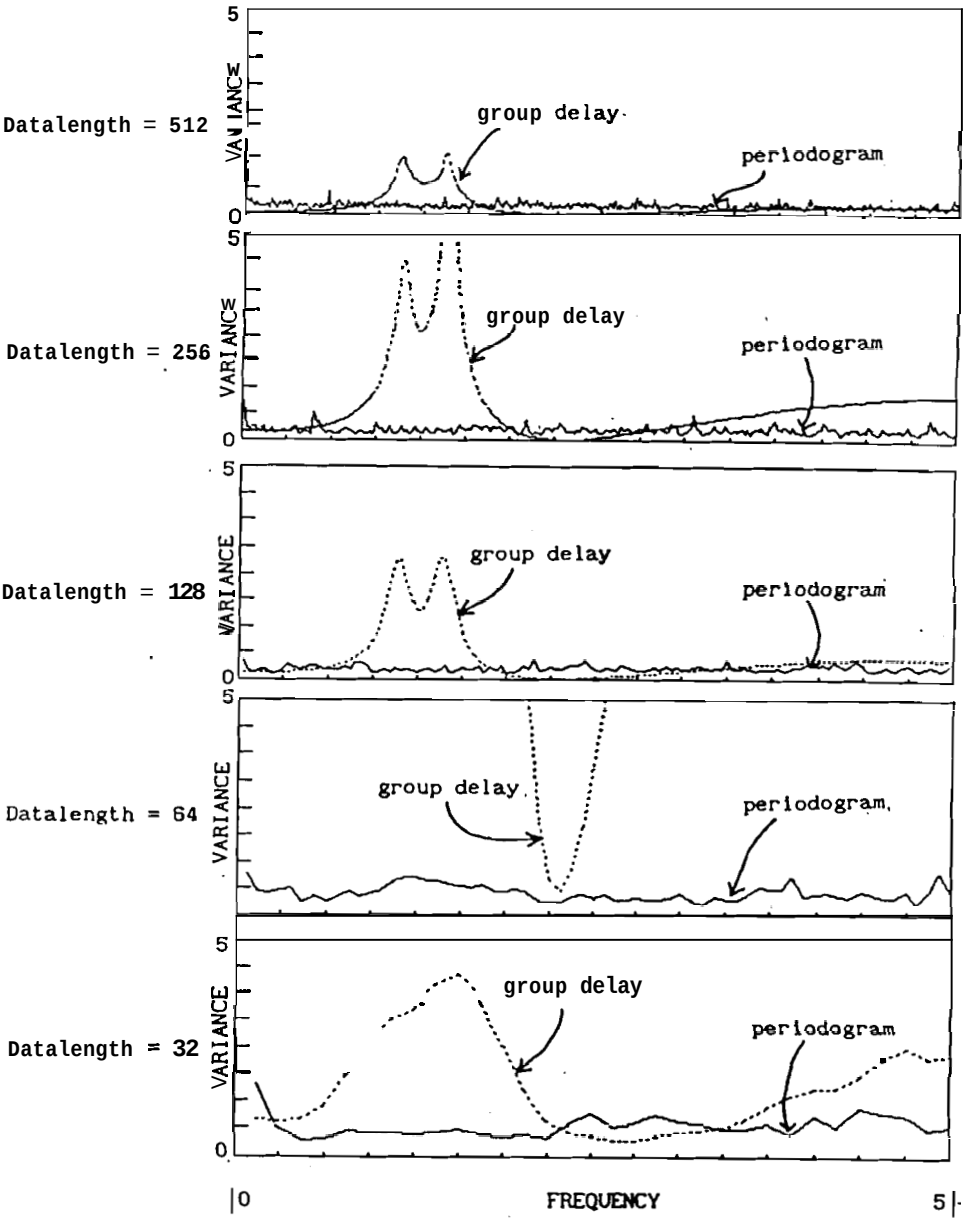


Fig.5.16. Illustration of the variance of estimates for AR process (clean) for different data lengths.

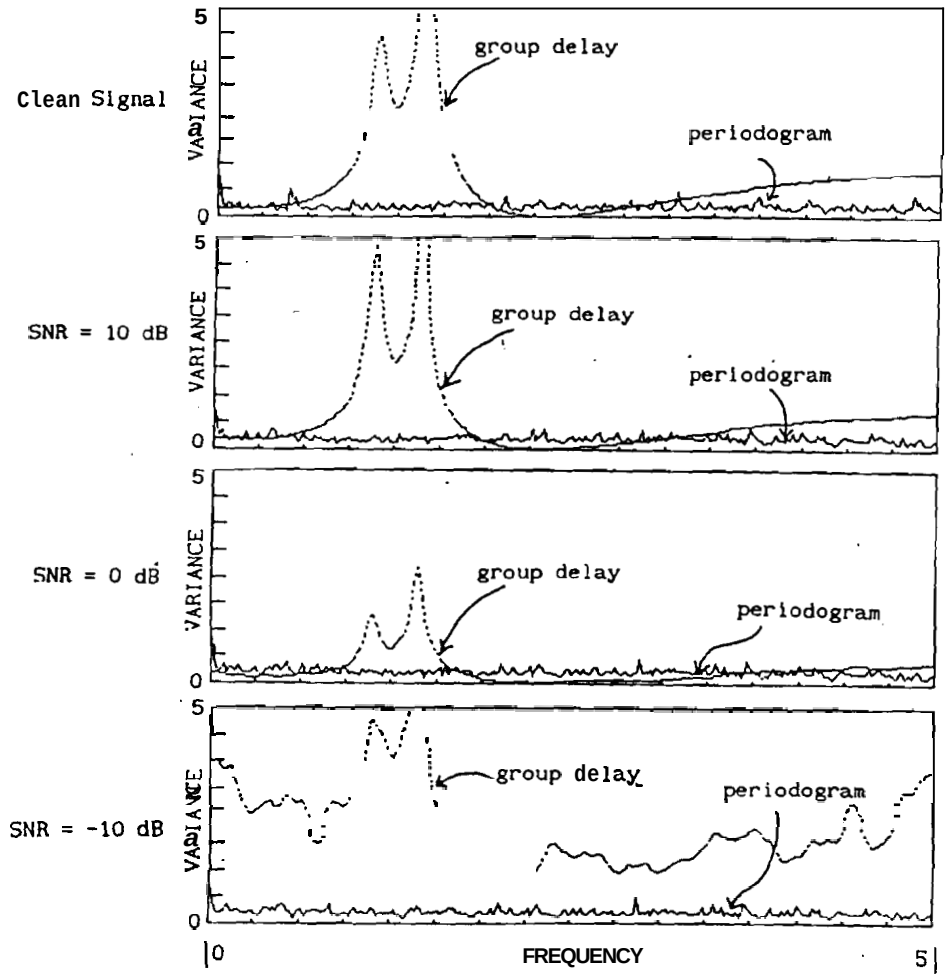


Fig.5.17. Illustration of the variance of estimates for AR process in noise for different noise levels.

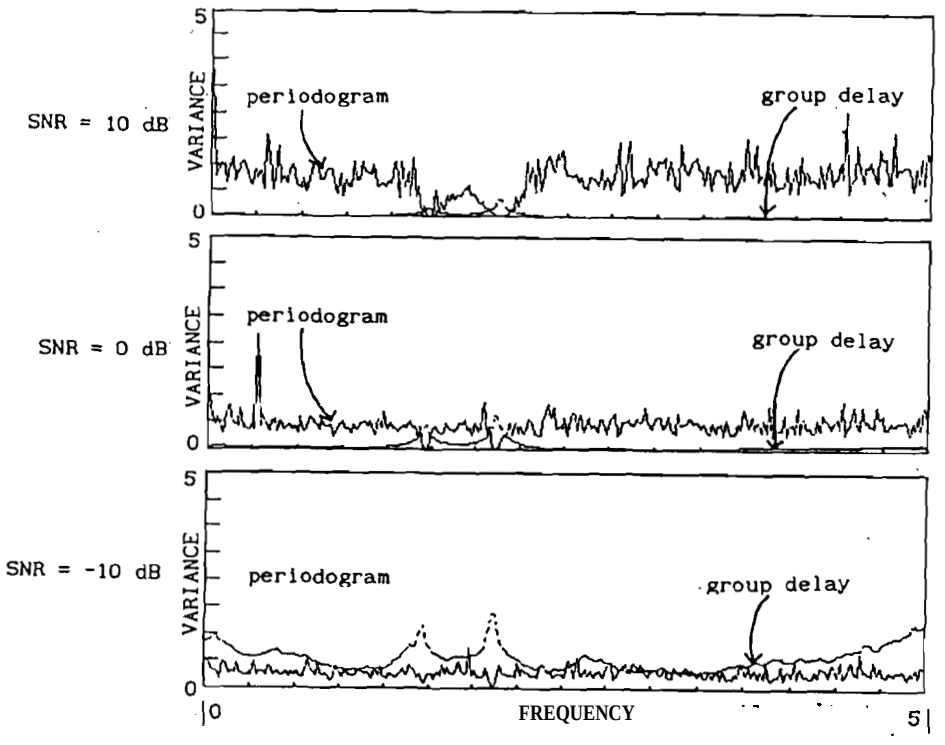


Fig.5.18. **Illustration of the variance of estimates for sinusoids in noise for different noise levels.**

sidelobes **(b)** has less effect on the bias, **(c)** restores the dynamic range and preserves the resolution of a periodogram estimate **(d)** works even for high noise levels and **(e)** performs better than model-based methods for noisy data, because resolution does not depend on factors like model order and spurious peaks are nearly absent even at high noise levels. However, comparison with model-based methods for short data records is not apt, because knowledge of the model definitely gives a better resolution than the periodogram estimate. Thus the proposed technique in its present form is not suitable for short data record analysis.

CHAPTER 6

SUMMARY AND CONCLUSIONS

6.1 Summary

The studies presented in this thesis represent an attempt to process the Fourier transform phase of signals for feature extraction.

Conventional methods for processing signals for parameter extraction rely heavily on the information that is available in the magnitude spectrum (or power spectrum (square of magnitude spectrum)) of the signal. This is because the features of a signal, for example periodicity manifests itself as picket fence harmonics in the magnitude spectrum of the signal, while they appear as phase transitions in the phase spectrum. But the phase spectrum of a signal is available only in wrapped form (restricted to the interval $\pm\pi$). The phase appears to be featureless and is hence difficult to interpret. If the phase spectrum of a signal is to be processed it should be first of all available in a unwrapped form. Although some algorithms are available for unwrapping the phase function, they are quite complex and do not work for all kinds of signals.

An alternative to processing the Fourier transform (FT) phase spectrum of the signal is processing the group delay function of the signal. The group delay function of the signal is defined as the negative derivative of the FT phase spectrum. The group delay function is easier to process (when compared to the phase spectrum) because it does not suffer from the wrapping problem and can be computed directly from the time domain signal. The focus of the research effort in this thesis is development of algorithms for processing the group delay function in a manner similar to that developed for processing the FT magnitude spectrum of a signal.

The group delay function suffers from the problem of poor sampling when the roots of the z-transform of the signal lie close to the unit circle in the z-domain. The poles of a signal that is generated as the output of a stable system are guaranteed to lie within the unit circle in the z-domain. But the zeros may lie within, on or outside the unit circle. To estimate parameters from the signal through group delay functions we must first of all be able to remove the zeros in the signal that lie on the unit circle. This is no trivial task as it is in general impossible to know the exact locations of the zeros apriori. Assuming a restricted task, namely estimating the parameters of the system in a source-system model for signal production (all-zero source and all-pole system), a group delay function with minimum phase characteristics is est'imated from the given signal. This is similar linear prediction analysis where minimum phase component of the spectrum is computed from the given signal. Application of this minimum phase group delay function for formant extraction is studied.

Another method for processsing the group delay function that is developed is the derivation of the modified group delay function directly from the standard group delay function. This approach is similar to the cepstrum analysis approach for processing the spectrum of a signal. In the modified group delay function approach to processing the group delay function of a signal, two different group delay functions are derived from the signal, one corresponding to that of the source and the other corresponding to that of the system. Application of these two functions for extraction of the model parameters in a source system model for speech, namely, pitch and formants is explored.

The modified group delay function has the property that the

zeros of a signal are suppressed, irrespective of whether the zeros are due to the source or not. Application of modified group delay function to process noisy speech is studied (as additive noise can be thought of as introduction of new zeros at random locations in the signal). Pitch and formant data are extracted from noisy speech and they are used used in a formant vocoder to synthesise speech.

The relationship between the group delay function of a minimum phase system and the spectrum is used to estimate the power spectrum of the signal from the modified group delay function. Application of the modified group delay function to estimate the spectra of random processes in noise is studied.

6.2 Major Contributions of the thesis

The most important contribution of this thesis is that it represents an attempt to process the phase spectrum of a signal for parameter extraction.

At first we are only able to estimate a minimum phase group delay function from the signal, but ultimately we arrive at an estimate of the power spectrum of the signal from the group delay function.

The strength of the proposed techniques seems to be that no model is forced on the signal. Therefore, the parameters estimated using the methods developed in this thesis should represent the underlying characteristics of the signal more accurately.

The, following algorithms are developed for speech analysis and spectrum estimation :

(a) A new algorithm for formant extraction from speech using a group delay function derived from the FT magnitude spectrum.

(b) A new algorithm for formant extraction from speech using a modified group delay function derived from FT phase.

(c) A new algorithm for pitch extraction from speech using the modified group delay function.

(d) A procedure to enhance speech using the data obtained in (b) and (c) for noisy speech.

(e) A new algorithm for spectrum estimation using modified group delay function.

6.3 *Criticisms of the work*

The major drawback of the work presented in this thesis seems to be the computational requirement. All the techniques developed are computationally expensive and therefore their use cannot be justified in practice.

Also most of the results presented are based on conjectures about the behaviour of signals. In most cases, we therefore substantiate our conjectures by simulation studies rather than through sound theoretical analyses.

To list some of the issues :

1. The proof of the fact that the signal derived from the magnitude spectrum in Chapter 3 is minimum phase is not established analyt'ically.

2. To estimate the modified group delay function in Chapter 4 a zero spectrum derived. It is observed that the approximate zero spectrum obtained is imperfect, in that the information corresponding to the resonances of signal is not completely suppressed in this spectrum (especially low frequency formants). Therefore, the regions corresponding to resonances in the modified group delay function are further emphasised by the remnant resonance information in the zero spectrum.

Therefore, estimation of an exact zero spectrum to suppress the zeros due excitation is still an issue.

3. In Chapter 5 most of the explanation are based on conjectures. No attempt is made to estimate the scale factor of the spectrum. Also analytical expressions for bias and variance are not given to evaluate the performance of the proposed method of spectrum estimation quantitatively.

6.4 Directions for future work

The work presented in this thesis uses a nonmodel based approach to processing the Fourier transform phase of signals. No attempt is model to phase spectrum corresponding to that of the system or source.

Although it may not be possible to model the wrapped phase, it may be possible to model the group delay function corresponding to that of the system. The advantage of a model-based approach to group delay processing of signals may be that group delay processing can be extended to analysis of short-data records.

APPENDIX A

A.1 Additive and High Resolution Property of Croup Delay functions:

Consider a causal, discrete time signal {x(n)} whose z-transform X(z) is a simple second order polynomial defined by

$$X(z) = (z - z_o^*)(z - z_o), \quad z_o = e^{-(\sigma_o + j\omega_o)} \tag{A.1}$$

and * indicates complex conjugation. $e^{-\sigma_o}$ determines the proximity of the zeros to the unit circle.

$$X(\omega) = X(z)\big|_{z=e^{j\omega}} = (e^{j\omega} - e^{-(\sigma_o + j\omega_o)})(e^{j\omega} - e^{-(\sigma_o - j\omega_o)}) \tag{A.2}$$

$$\theta(\omega) = \tan^{-1}\left[\frac{\sin\omega - e^{-\sigma_o}\sin\omega_o}{\cos\omega - e^{-\sigma_o}\cos\omega_o}\right] + \tan^{-1}\left[\frac{\sin\omega + e^{-\sigma_o}\sin\omega_o}{\cos\omega - e^{-\sigma_o}\cos\omega_o}\right] \tag{A.3}$$

Using the rules of differentiation we can show that

$$\theta'(\omega) = \theta_1'(\omega) + \theta_2'(\omega)$$

where $\theta_1'(\omega)$ is the phase corresponding to that of the first term in eq(A.3) and $\theta_2'(\omega)$ is the phase corresponding to that of the second term in eq(A.3). Defining

$$\tau(\omega) = -\theta'(\omega)$$

it follows that

$$\tau(\omega) = \tau_1(\omega) + \tau_2(\omega)$$

where $\tau(\omega)$ is the overall group delay function and $\tau_1(\omega)$ and $\tau_2(\omega)$ are the group delay functions of the component group delay functions corresponding to the complex conjugate zero pair. Using eq(A.3) $\tau(\omega)$ can be obtained as

$$\tau(\omega) = -\left[\frac{1 - e^{-\sigma_o}\cos(\omega-\omega_o)}{1 + e^{-2\sigma_o} - 2e^{-\sigma_o}\cos(\omega-\omega_o)}\right] - \left[\frac{1 - e^{-\sigma_o}\cos(\omega+\omega_o)}{1 + e^{-2\sigma_o} - 2e^{-\sigma_o}\cos(\omega+\omega_o)}\right] \tag{A.4}$$

Consider

$$\tau_1(\omega) = -\left[\frac{1 - e^{-\sigma_o}\cos(\omega-\omega_o)}{1 + e^{-2\sigma_o} - 2e^{-\sigma_o}\cos(\omega-\omega_o)}\right]$$

Equating $\tau_1'(\omega) = 0$ we get

$$(e^{-\sigma_o} - e^{-3\sigma_o})\sin(\omega-\omega_o) = 0 \qquad (A.5)$$

$(e^{-\sigma_o} - e^{-3\sigma_o})$ can never be zero unless $\sigma = 0$. Therefore in general eq(A.5) becomes zero when $w-w_\Theta = 0, \pi, 2\pi$ .... Taking the derivative of eq(A.5) we get

$$\tau_1''(\omega)\big|_{\omega=\omega_o} = (e^{-\sigma_o} - e^{-3\sigma_o}) \qquad (A.6)$$

Eq(A.6) = 0 when $\sigma_\Theta = 0$.

Eq(A.6) is negative when $e^{-3\sigma_o} > e^{-\sigma_o}$.

Eq(A.6) is positive when $e^{-3\sigma_o} < e^{-\sigma_o}$

When the root is exactly on the unit circle, $\sigma_\Theta = 0$.

When the root is inside the unit circle, $e^{-\sigma_o} < 1$, $e^{-3\sigma_o} < e^{-\sigma_o}$ and the above expresssion is positive and $\tau_1(\omega)$ goes through a minimum at $\omega=\omega_o$.

When the root is outside the unit circle, $e^{-\sigma_o} > 1$, $e^{-3\sigma_o} > e^{-\sigma_o}$ and hence the above is negative and $\tau_1(\omega)$ goes through a maximum at $\omega=\omega_\Theta$.

The same argument applies for $\tau_2(\omega)$ also. When X(z) is defined by the reciprocal of the second order system defined in (A.1), the corresponding group delay function is the negative that derived in eq(A.4). Fig.A.1 is an illustration of the group delay function for different first and second order polynomials. The dotted curves correspond to the group delay function of poles at the same locations as the zeros.

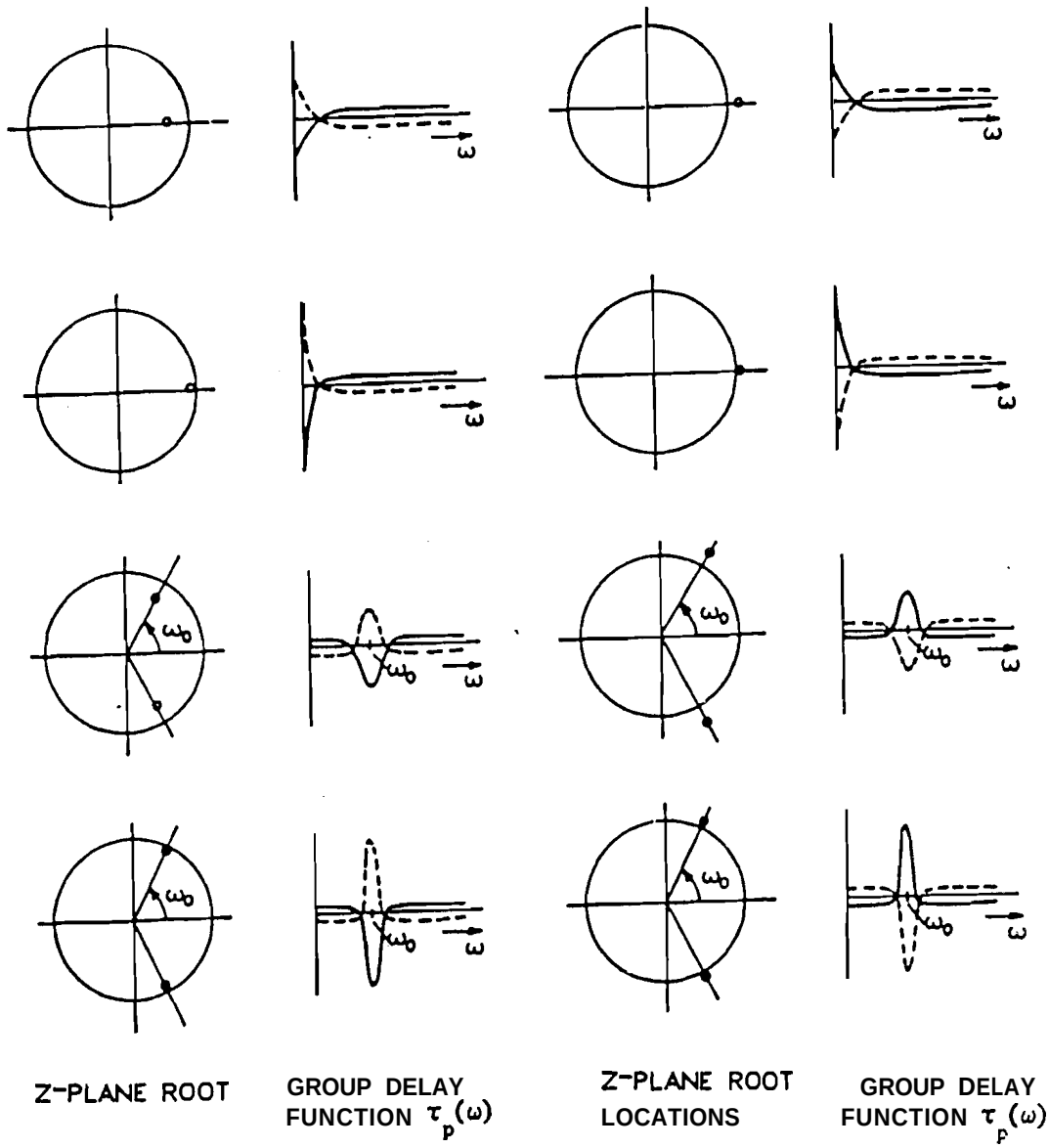


Fig. A. 1 Illustration of the group delay functions for different first- and second-order polynomials. The dotted curves correspond to poles in the z-plane at the same locations as the zeros.

APPENDIX B

B.1 Modified Group Delay function :

Consider a signal x(n) whose z-transform is given **by**

$$X(z) = \frac{1 - \gamma z^{-T_o}}{(z - z_o)(z - z_o^*)} \qquad (B.1)$$

where z_Θ is given by $z_o = e^{-(\sigma_o + j\omega_o)}$. This system corresponds to the excitation of a system which consists of two poles in complex conjugate locations by two impulses separated **by** a period T_Θ. The amplitude of the first impulse is unity while the amplitude of the second impulse is γ

Now

$$\tau(\omega) = -\left[\frac{\gamma T_o \cos\omega T_o - \gamma^2 T_o}{1 + \gamma^2 - 2\gamma\cos\omega T_o}\right] + \left[\frac{1 - e^{-\sigma_o}\cos(\omega-\omega_o)}{1 + e^{-2\sigma_o} - 2e^{-\sigma_o}\cos(\omega-\omega_o)}\right]$$

$$+ \left[\frac{1 - e^{-\sigma_o}\cos(\omega+\omega_o)}{1 + e^{-2\sigma_o} - 2e^{-\sigma_o}\cos(\omega+\omega_o)}\right]. \qquad (B.2)$$

The modified group delay function $\tau_\Theta(\omega)$ is defined as

$$\tau_\Theta(\omega) = \tau(\omega).(1 + \gamma^2 - 2\gamma\cos\omega T_o) \qquad (B.3)$$

Consider the term

$$\tau_{o1}(\omega) = -\left[\gamma T_o \cos\omega T_o - \gamma^2 T_o\right] +$$

$$\left[\frac{1 - e^{-\sigma_o}\cos(\omega-\omega_o)}{1 + e^{-2\sigma_o} - 2e^{-\sigma_o}\cos(\omega-\omega_o)}\right].\left[1 + \gamma^2 - 2\gamma\cos\omega T_o\right] \qquad (B.4)$$

Computing $\tau'_{\Theta 1}(\omega)$ we get

$$\tau'_{\Theta 1}(W) = -\left[-2\gamma\cos\omega T + \gamma + 1)\right].$$

$$\left[\frac{2e^{-\sigma_o}(1 - e^{-\sigma_o}\cos(\omega-\omega_o))\sin(\omega-\omega_o)}{(-2e^{-\sigma_o}\cos(\omega-\omega) + e^{-2\sigma_o} + 1^2}\right.$$

$$+ \frac{e^{-\sigma_o}\sin(\omega-\omega_o)}{(-2e^{-\sigma_o}\cos(\omega-\omega_o) + e^{-2\sigma_o} + 1)} \Bigg]$$

$$- 2\gamma\sin\omega T_o \left[\frac{1 - e^{-\sigma_o}\cos(\omega-\omega_o)}{(-2e^{-\sigma_o}\cos(\omega-\omega_o) + e^{-2\sigma_o} + 1)}\right] T_o$$

$$- \gamma T_o^2 \sin\omega T_o .$$

$$\tau''_{o1}(\omega)\Big|_{\substack{\omega=\omega_o \\ \omega T_o = \pm(2n+1)\pi}} = -(1-\gamma)^2 \frac{e^{-\sigma_o}}{(1-e^{-\sigma_o})^2}\left[\frac{(1+e^{-\sigma_o})}{(1-e^{-\sigma_o})}\right]$$

$$+\gamma T_o^2 \left[\frac{2}{(1-e^{-\sigma_o})^2} - T_o\right]$$

τ''_{o1} is negative as the second term will be larger than the first term since both $e^{-\sigma_o}$ (since we are considering a stable system) and γ will be utmost 1, the second term is negative since T_o is in samples ($\gamma T_o^3 > \alpha\gamma T_o^2$, a $\ll$ T , is some constant). Therefore, $\tau_{o1}(\omega)$ is a maximum at $\omega=\omega_o$, provided $\cos\omega T_o = -1$. This essentially means that the antiresonance frequency of a zero ***must not coincide*** with resonance frequency of a pole.

REFERENCES

W.A. Ainsworth (19881, ***Speech Recognition by Machine,*** Peter Peregrinus Ltd., on behalf of the Institution of Electrical Engineers.

M.S. Andrews, J. Picone and R.D. Degroat, "Robust Pitch Determination via SVD Based Cepstral Methods", ***Proc. ICASSP-90,*** pp.253-256.

B.S. Atal and S.L. Hanauer (19711, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", J. ***Acoust. Soc. Am.,*** Vol.50, pp.637-655.

B.S. Atal and L.R. Rabiner (19761, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition", IEEE ***Trans. Acoust. Speech and Signal Process.,*** Vol. ASSP-24, No.3, pp. 201-212.

B.S. Atal and M.R. Schroeder (1979), "Predictive Coding of Speech Signals and Subjective Error Criteria", IEEE ***Trans. Acoust. Speech and Signal Process.***, Vol. ASSP-27, No.3, pp.247-253.

R.N. Bracewell (19861, ***The Fourier Transform and its Applications,*** McGraw Hill, New York.

A.J. Berkhout (1973), "On the Minimum-Length Property of One-Sided Signals", ***Geophysics,*** Vol.38, No.4, pp.657-672.

A.J. Berkhout (1974), "Related Properties of Minimum-Phase and Zero-Phase Time Functions", ***Geophysical Prospecting,*** Vol.22, pp.683-709.

S.F. Boll (19791, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE ***Trans. Acoust. Speech and Signal Process.***, Vol. ASSP-27, No.2, pp.113-120.

S.F. Boll and D.C. Pulsipher (19801, "Suppression of Acoustic Noise in Speech Using Two Microphone Noise Cancellation", IEEE ***Trans. Acoust. Speech and Signal Process.***, Vol. ASSP-28, No.6, pp.752-753.

H. Chhatwal and A.G. Constantinides (1987), "Speech Spectral Segmentation for Spectral Estimation and Formant Modeling", *Proc. **ICASSP-87,*** pp.316-319.

D.G. Childers, D.P. Skinner and R.C. Kemerait (1977), "The Cepstrum: A Guide to Processing", ***Proc.*** IEEE, Vol.65, No.10, pp.1428-1442.

D.G. Childers (editor) (1978), ***Modern Spectrum Estimation,*** IEEE Press.

S.L. Curtis and A.V. Oppenheim (1987), "Reconstruction of Multidimensional Signals from Zero Crossings", J. ***Opt. Soc. of Am.,*** Vol.4, No.1, pp.22i-231.

E. Denoel and J.-P. Solvay (1985), "Linear Prediction Speech With Least Absolute Error Criterion", IEEE ***Trans. Acoust. Speech and Signal Process.***, Vol. ASSP-33, No.6, pp.1397-1403.

G. Duncan and M. A. Jack, "Pole Focusing: A New Approach to LPC Analysis offering Superior Noise Robustness and Feature Resolution", J. *IETE* **Spl. *Issue* on Speech Process.**, Vol.34, No.1, pp.29-37.

H. K. Dunn (19611, "Methods of Measuring Vowel Formant Bandwidths", J. **Acoust. *Soc.* Am.**, Vol.33, pp.1737-1746.

G. Fant, "The Acoustics of Speech", **Proc. Third Znt. Cong. on Acoust.**, pp.188-201, 1959.

A. El-Jaroudi and J. Makhoul (19871, "Discrete All-Pole Modeling of Voiced Speech", **Proc. ZCASSP-87,** pp.320-323.

J. L. Flanagan (19561, "Automatic Extraction of Formant Frequencies from Continuous Speech", J. **Acoust. *Soc.* Am., Vol.28,** pp.110-118.

J. L. Flanagan and L. Cherry (1969), "Excitation of Vocal tract Synthesisers", J. **Acoust. *Soc.* Am., Vol.45,** pp.764-769.

J. L. Flanagan (1972), "Voices and Machines", J. **Acoust. *Soc.* Am.,** V01.5, pp.1375-1387.

G. Fant (1959), "The Acoustics of Speech", **Proc. Third Znt. Cong. on Acoust..** pp.181-201.

J. R. Fienup (19871, "Reconstruction of a complex-valued object from the modulus of the Fourier transform using a support constraint", J. **Opt. *Soc.* Am,** Vol.4, No.1, pp.118-123.

R. H. Frazier, S. Samsam, L. D. Braida and A. V. Oppenheim (19761, "Enhancement of Speech by Adaptive Filtering", **Proc. ZCASSP-76,** pp.251-253.

D. G. Ghiglia, Gary A. Mastin and Louis A. Romero, "Cellular Automata method for phase unwrapping", J. **Opt. *Soc.* Am.,** Vol.4, No.1, pp. 267-280.

Y. Gong and J. P. Haton (19871, "Time Domain Harmonic Matching Pitch Estimation Using Time-Dependent Speech Modeling". **ZEEE Trans. Acoust. Speech and Signal Process.,** Vol. ASSP-35, No.10, pp.1386-1400.

Hema A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana (1989a), "Formant extraction from Fourier transform phase," **Proc. ZCASSP-89,** Glasgow, pp.484-487.

Hema A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana (1989b), "Formant extraction from phase using weighted group delay function," Electronics letters, 9th Nov., Vol.25, No.23, pp.1609-1611.

M. H. Hayes, J. S. Lim and A. V. Oppenheim (19801, "Signal Reconstruction from Phase or Magnitude", **ZEEE Trans. Acoust. Speech and Signal Process.**, Vol. ASSP-28, No.6, pp.672-680.

S. S. Haykin (1986), **Adaptive Filter Theory,** Prentice-Hall EngleWood Cliffs :New Jersey.

L. Hodgson, M. E. Jernigan and B. W. Wills, "Nonlinear Multiplicative Cepstral Analysis for Pitch Extraction from Speech", **Proc. ZCASSP-90,**

pp.257-260

F.Itakura and S.Salto (19701, "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies", ***Electron. and Commn.***, Vol.53-A, pp.36-43.

S.M.Kay (19881, ***Modern Spectrum Estimation: Theory and Application,*** Prentice Hall Englewood Cliffs : New Jersey.

S.B.Kesler (editor) (19861, ***Modern Spectrum Estimation,*** *II,* IEEE Press.

G.E.Kopec (1986a), "A Family of Formant Trackers Based on Hidden Markov Models", ***Proc. ICASSP-86,*** pp.1225-1228.

G.E.Kopec (1986b), "Formant Tracking Using Hidden Markov Models and Vector Quantization", ***IEEE Trans. Acoust. Speech and Signal Process.,*** Vol.ASSP-34, No.4, pp.709-729.

C.H.Lee (19871, "Robust Linear Prediction for Speech Analysis", ***Proc. ICASSP-87,*** pp.289-282.

C.H.Lee (19881, "On Robust Linear Prediction", ***IEEE Trans. Acoust. Speech and Signal Process.***, Vol.36, No.5, pp.642-640.

J.S.Lim and A.V.Oppenheim (1978a), "All-Pole Modeling of Degraded Speech", ***IEEE Trans. Acoust. Speech and Signal Process.***, Vol.ASSP-26, No.3, pp.197-210.

J.S.Lim, A.V.Oppenheim and L.D.Braida (1978b), "Evaluation of an Adaptive Combing Filtering Method for Enhancing Speech Degraded by White Noise Addition", ***IEEE Trans. Acoust. Speech and Signal Process.,*** Vol.ASSP-26, No.5, pp.354-358.

J.S.Lim (1979a), "Spectral Root Homomorphic Deconvolution System", ***IEEE Trans. Acoust. Signal Process.,*** Vol.ASSP-27, pp.223-231.

J.S.Lim (1979b), "Enhancement and Bandwidth Compression of Noisy Speech", ***Proc. IEEE*** Vol.67, No.12, pp.1586-1604.

K.V.Madhu Murthy and B.Yegnanarayana (19891, "Effectiveness of Representation of Signals through Group Delay Functions", ***Signal Processing,*** 17, 141-150, Elsevier Science Publishers, B.V.

J.Makhoul (19751, "Linear Prediction: A Tutorial Review", ***Proc. IEEE,*** Vol.63, No.4, pp.561-580.

D.Malah (1979), "Time-Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals", ***IEEE Trans. Acoust. Signal Process.,*** Vol.ASSP-27, No.2, pp.121-133.

M.T.Manry (1985), "Signal Processing using Implicit Phase", ***IEEE Trans. Circuits and Systems,*** Vol.CAS-32, No.2, pp.150-159.

C.P.Mariadassou and B.Yegnanarayana (19901, "Image Reconstruction from Noisy Holograms", *IEE-Proc.(London),* Vol.137, Part F, No.5.

J.D.Markel (1972a), "Digital Inverse Filtering - A New Tool for

Formant Trajectory Estimation", ***IEEE Trans. Audio Electroacoust.,*** Vol. AU-20, No.3, pp.129-137.

J.D. Markel (1972b), "The SIFT Algorithm for Fundamental Frequency Estimation", ***IEEE Trans. Audio Electroacoust.,*** Vol. AU-20, No.6, pp.367-377.

S.L. Marple (19871. ***Digital Spectral Analysis with Applications,*** Prentice-Hall, Englewood Cliffs : New Jersey.

S.S. McCandless (19741, "An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra", ***IEEE Trans. Acoust. Speech and Signal Process.,*** Vol. ASSP-22, No.2, pp.135-141.

R.J. McAulay (1984), "Maximum Likelihood Spectral Estimation and its Application to Narrow-Band Speech Coding", ***IEEE Trans. Acoust. Speech and Signal Process.,*** Vol. ASSP-32, No.2, pp.243-251.

R.J. McAulay and T.F. Quatieri (19861, "Speech Analysis/Synthesis Based on a Sinusoidal Representation", ***IEEE Trans. Acoust. Speech and Signal Process.,*** Vol. ASSP-34, No.4, pp.744-754.

N.J. Miller (1975), "Pitch Detection by Data Reduction", ***IEEE Trans. Acoust. Speech and Signal Process.*** (Special Issue on IEEE Symposium on Speech Recognition), Vol. ASSP-23, No.1, pp.72-79.

H. Morikawa and H. Fujisaki (19841, "System Identification of the Speech Production Process Based on a State-Space Representation", Vol. ASSP-32, No.2, pp.252-262.

H.A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana (1989a), "Formant Extraction from Fourier Transform Phase", ***Proc. ICASSP-89,*** Clasgow, U.K., pp.484-487.

H.A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana (1989b), "Formant Extraction from Phase Using Weighted Croup Delay Function", ***Electronics Letters,*** Vol. 25, No.23, pp.1609-1611.

A.M. Noll (1967), "Cepstrum Pitch Determination", J. ***Acoust. Soc. Am.,*** Vol. 41, No.1, pp.293-309.

A.V. Oppenheim and R.W. Schafer (19631, "Homomorphic Analysis of Speech", ***IEEE Trans. Audio Electroacoust.,*** Vol. AU-16, No.3, pp.221-226.

A.V. Oppenheim and R.W. Schafer (1975), ***Digital Signal Processing,*** EngleWood Cliffs, New Jersey: Prentice-Hall.

A.V. Oppenheim and J.S. Lim (1981), "The Importance Phase in Signals", ***Proc. IEEE,*** Vol. 69, No.3, pp.529-541.

A. Papoulis (1977), ***Signal Analysis,*** McGraw Hill, New York.

E.N. Pinson (1963), "Pitch Synchronous Time-Domain Estimation of Formant Frequencies and Bandwidths", J. ***Acoust. Soc. Am.,*** Vol. 35, pp.1264-1273.

M.R. Portnoff (1981a), "Short-Time Fourier Analysis of Sampled Speech",

IEEE Trans. Acoust. Speech and Signal Process., Vol.ASSP-29, No.3, pp.364-373.

M.R.Portnoff (1981b),"Time-Scale Modification of Speech Based on Short-Time Fourier Analysis", *IEEE Trans. Acoust. Speech and Signal Process.*, Vol.ASSP-29, No.3, pp.374-390.

T.F.Quatieri, Jr. (1979), "Minimum and Mixed Phase Speech Analysis-Synthesis by Adaptive Homomorphic Deconvolution", *IEEE Trans. Acoust. Speech and Signal Process.*, Vol.ASSP-27, No.4, pp.328-335.

L.R.Rabiner, M.J.Cheng, A.E.Rosenberg and C.A.McGonegal (1976), "A Comparative Performance Study of Several Pitch Detection Algorithms", *IEEE Trans. Acoust. Speech and Signal Process.*, Vol.ASSP-24, No.5, pp.399-417.

L.R.Rabiner and R.W. Schafer (1978), *Digital Processing of Speech Signals*, EngleWood Cliffs, New Jersey:Prentice-Hall.

N.S.Reddy and M.N.S.Swamy (1985), "Derivative of Phase Spectrum of Truncated Autoregressive Signals", *IEEE Trans. Circuits and Systems*, Vol.CAS-32, No.6, pp.616-618.

G.Regoll (1986), "A New Algorithm for Estimation of Formant trajectories Directly from the Speech Signal Based on an Extended Kalman Filter", *Proc. ICASSP-86*, pp.1225-1228.

M.J.Ross, H.L.Shaffer, A.Cohen, R.Freudberg and H.J.Minley, (1974) "Average Magnitude Difference Function Pitch Extraction", *IEEE Trans. Acoust. Speech, Signal Process.*, Vol.ASSP-22, No.5, pp.353-362.

M.R.Sambur (1978), "Adaptive Noise Cancelling for Speech Signals", *IEEE Trans. Acoust. Speech and Signal Process.*, Vol.ASSP-26, pp.419-423.

R.W.Schafer and L.R.Rabiner (19701, "System for Automatic Formant Analysis of Voiced Speech", J. *Acoust. Soc. Am.*, Vol.47, pp.634-648.

R.W.Schafer and L.R.Rabiner (1975), "Digital Representation of Speech signals", *Proc. IEEE*, Vol.63, pp.662-677.

J.Schroeter, J.N.Larar and M.M.Sondhi (1987), "Speech Parameter Estimation Using a Vocal Tract/Cord Model", *Proc. ICASSP-87*, pp.308-311.

M.Slaney (19901, "A Perceptual Pitch Detector", *Proc. ICASSP-90*, pp.357-360.

K.H.Song and C.K.Un (19831, "Pole-Zero Modeling of Speech Based on High-Order Pole-Model Fitting and Decomposition Method", *IEEE Trans. Acoust. Speech and Signal Process.*, Vol.ASSP-31, No.6, pp.1556-1565.

M.M.Sondhi (19681, "New Methods of Pitch Extraction", *IEEE Trans.*

Audio Electroacoust., Vol.Au-16, No.3, pp.262-266.

J.M.Tribolet (1979), "A New Phase Unwrapping Algorithm", ***IEEE Trans. Acoust. Speech and Signal Process.***, Vol.ASSP-, No.2, pp.170-179.

P.L.Van Hove, M.H.Hayes, J.S.Lim and A.V.Oppenheim (1983), "Signal Reconstruction from Signed Fourier transform Magnitude", ***IEEE Trans. Acoust. Speech and Signal Process.***, Vol.ASSP-31, No.5, pp.1286-1292.

W.Verhelst and O.Steenhaut (1986), "A New Model for the Short-Time Cepstrum of Voiced Speech", ***IEEE Trans. Acoust. Speech and Signal Process.***, Vol.ASSP-34, No.1, pp.43-51.

B.Widrow and Stearns (1986), ***Adaptive Signal Processing,*** EngleWood Cliffs, Prentice-Hall:New Jersey.

B.Yegnanarayana (1978), "Formant Extraction from Linear Prediction Phase Spectra", J. ***Acoust.*** *Soc.* ***Am.,*** Vol.63, pp.1638-1640.

B.Yegnanarayana (1981), "Design of ARMA Digital Filters by Pole-Zero Decomposition", ***IEEE Trans. Acoust. Speech and Signal Process.,*** Vol.ASSP-29, No.3, pp.433-439.

B.Yegnanarayana (1981), "Speech Analysis by Pole-Zero Decomposition of Short Time Spectra", ***Signal Process.*** pp. 5-17.

B.Yegnanarayana, D.K.Saikia and T.R.Krishnan (1984), "Significance of Group Delay Functions in Signal Reconstruction from Spectral Magnitude or Phase", ***IEEE Trans. Acoust. Speech and Signal Process.,*** Vol.ASSP-32, No.3, pp.610-623.

B.Yegnanarayana, J.Sreekanth and A. Rangarajan (1985), "Waveform Estimation Using Group Delay Processing", ***IEEE Trans. Acoust. Speech and Signal Process.,*** Vol.ASSP-33, No.4, pp.832-836.

B.Yegnanarayana, S.T.Fathima and H.A.Murthy (1987a), "Reconstruction from Fourier Transform Phase with Applications to Speech Analysis", ***Proc. ICASSP-87,*** Dallas, Texas, pp. 301-304.

B.Yegnanarayana, C.P.Mariadassou and Pramod Saini (1990), "Signal Reconstruction from Partial Data for Sensor Array Imaging Applications", ***Signal Processing,*** Vol.19, No.2.

LIST OF PUBLICATIONS

Publications *(in refereed Foreign Journals)* :

(1) Hema A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana, "Formant extraction from phase using weighted group delay function," ***Electronics letters,*** 9th Nov., 1989, Vol.25, No.23, pp.1609-1611.

(2) Hema A. Murthy and B. Yegnanarayana. "Speech processing using Group delay functions," ***Signal Processing,*** Vol.22, pp.259-267, March, 1991.

(3) Hema A. Murthy and B. Yegnanarayana "Formant extraction from group delay function," ***Speech Communication,*** Vol.10, pp.209-221, August, 1991.

(4) B.Yegnanarayana and Hema A. Murthy, "Significance of Group Delay functions in Spectrum estimation," to appear in Sept 1992 issue of ***IEEE Trans. on Signal Processing.***

Publications *(in Indian Journals)* :

(1) B.Yegnanarayana, K.V. Madhu Murthy and Hema A. Murthy, "Applications of Group Delay Functions in speech processing," Invited Paper, ***JIETE Special issue on Speech processing,*** pp.20-29, Jan 1988.

hblications *(in International Conference Proceedings)* :

(1) B. Yegnanarayana, S. T. Fathima and Hema A. Murthy, "Signal reconstruction from Fourier Transform phase with applications to speechanalysis," ***ICASSP-87,*** Dallas, Texas.

(2) B. Yegnanarayana, K.V. Madhu Murthy and Hema A. Murthy, "Processing of noisy speech using partial phase," ***European Conference on Speech Technology,*** Edinborough, U.K., Vol.1, pp.203-206, Sept., 1987.

(3) B.Yegnanarayana George E. Duncan and Hema A. Murthy, "Improving feature extraction from speech using minimum phase group delay spectra," ***EUSIPCO-88,*** Grenoble, France, Vol.1, pp.267-270, Sept', 1988.

(4) Hema A. Murthy, A. A. Babu, B.Yegnanarayana and K.V. Madhu Murthy", Speech coding using Fourier Transform phase," ***EUSIPCO-88,*** Grenoble, France, Vol.2, pp.879-882, Sept., 1988.

(5) Hema A. Murthy, K.V. Madhu Murthy and B. Yegnanarayana, "Formant extraction from Fourier transform phase," ***ICASSP-89,*** Glasgow, Scotland, Vo1.1, pp.484-487, May, 1989.

(6) G.Duncan, B.Yegnanarayana and Hema A. Murthy, "A nonparametric method of formant estimation using group delay spectra," ***ICASSP-89,*** Glasgow, Scotland, Vol.1, pp.572-575, May, 1989.

(7) B.Yegnanarayana, Hema A. Murthy and V.R.Ramachandran, "Speech Enhancement using group delay functions," ***ICSLP-90,*** Kobe, Japan, Vol.2, pp.301-304, Nov., 1990.

(8) B. Yegnanarayana and Hema A. Murthy. "New Methods of processing noisy signals using group delay functions." ICCS-90, Singapore, Nov., 1990.

(9) B. Yegnanarayana, Hema A. Murthy, R. Sundar, V. R. Ramachandran, A. S. Madhukumar, N. Alwar and S. Rajendran, "Development of a text-to-speech system for Indian Languages," KBCS-90, Pune, India, Dec., 1990.

(10) B. Yegnanarayana, Hema A. Murthy and V. R. Ramachandran. "Processing of noisy speech using modified group delay functions", ICASSP-91, Toronto, Canada, pp.945-948, May, 1991.

Publications(in National Conference Proceedings):

(1) S. R. Rajesh Kumar, V. R. Ramachandran, A. S. Madhukumar, Hema A. Murthy and B. Yegnanarayana, "A text-to-speech system for Indian languages," presented at the Seminar on Common Phonetic Matrix for Indian Languages, CIIL, Mysore, India, March, 1990.

(2) B. Yegnanarayana and Hema A. Murthy, "Spectrum Estimation using Fourier transform Phase," Proc. of the workshop on Signals, Systems, Communications and Networking, Bangalore, India, pp.44-53, July, 1990.