TIME-VARYING LINEAR PREDICTIVE CODING OF SPEECH SIGNALS

by

MARK GILBERT HALL

B.S., B.A., Nebraska Wesleyan University
(1973)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August, 1977

Signature of Author...
Department of Electrical Engineering and Computer Science,
August 26, 1977

Certified by...
Thesis Supervisor

Accepted by...
Chairman, Departmental Committee on Graduate Students

TIME-VARYING LINEAR PREDICTIVE CODING OF SPEECH SIGNALS

by

MARK GILBERT HALL

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the Degree of Master of Science.

ABSTRACT

For linear predictive coding (LPC) of speech, the speech waveform is modelled as the output of an all-pole filter. The waveform is divided into many short intervals (10-30 msec) during which the speech signal is assumed to be stationary. For each interval the constant coefficients of the all-pole filter are estimated by linear prediction by minimizing a squared prediction error criterion. This thesis investigates a modification of LPC, called time-varying LPC, which can be used to analyze nonstationary speech signals. In this method, each coefficient of the all-pole filter is allowed to be time-varying by assuming it is a linear combination of a set of known time functions. The coefficients of the linear combination of functions are obtained by the same least squares error technique used by the LPC. Methods are developed for measuring and assessing the performance of time-varying LPC and results are given from the time-varying LPC analysis of both synthetic and real speech.

Thesis Supervisor: Alan S. Willsky

Title: Associate Professor of Electrical Engineering and Computer Science

ACKNOWLEDGEMENTS

I cannot adequately express my appreciation for the guidance and support given to me during this research by both my thesis supervisor, Prof. Alan Willsky and by Prof. Alan Oppenheim. Their contributions to this thesis through their advice and many suggestions were invaluable.

I would like to thank Jae Lim for the much needed assistance he gave while I was learning how to use the computer facility.

I would also like to thank Debi Lauricella, who typed a beautiful document in spite of my handwriting and constant changes.

The financial support of the Naval Surface Weapons Center, Dahlgren, Va., was very much appreciated during this year of graduate study.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

There are many applications which involve the processing and representation of speech signals [1]. One class of applications is concerned with the analysis of the speech waveform. Some examples which use speech analysis include speaker verification or identification, speech recognition, and the acoustic analysis of speech. Another area of interest is in the synthesis of speech, which could be used for automatic reading machines or for creating a voice response from a computer. A third type of application involves both the analysis and synthesis of speech. An example of this would be the data rate compression used for the efficient coding of speech for transmission and reproduction.

Many different techniques and models can be implemented for these applications. One method, that is based on the structure of the speech waveform, represents the physical speech production system as a slowly time-varying linear system which is excited by an appropriate input signal [1,3,16].

To illustrate why this is a reasonable model, an acoustical waveform is shown in figure 1.1a. It is evident from the signal that even though the general characteristics of the waveform are changing with time, there are segments where the form of the signal remains relatively constant. These segments can be

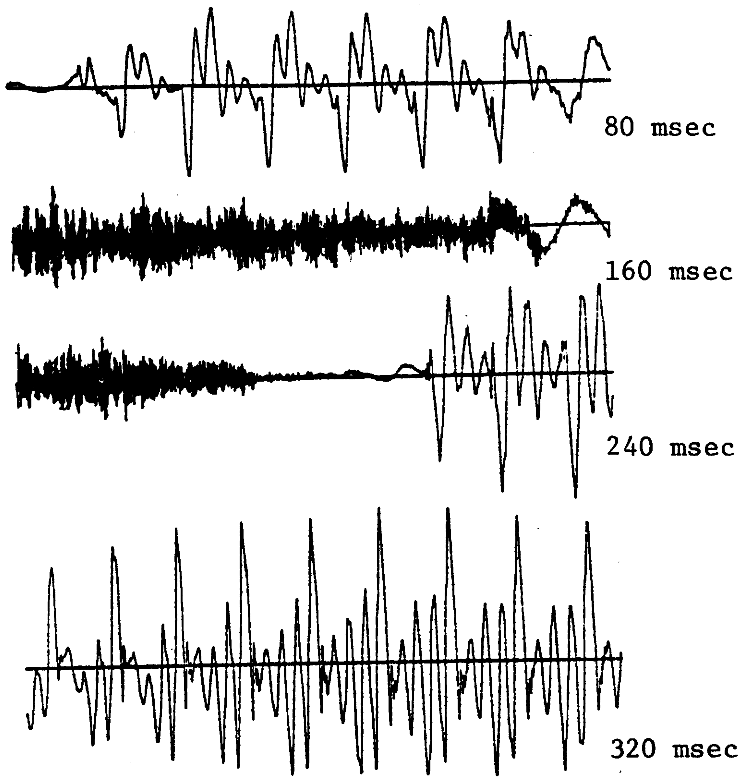

Figure 1.1a Speech Waveform Example

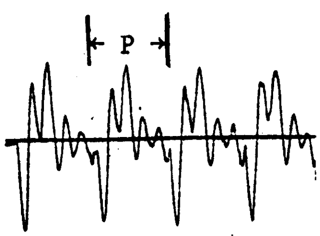

Figure 1.1b Voiced Sound
Pitch Period - P

Figure 1.1c Unvoiced Sound

classified as voiced or unvoiced.

The voiced segments are nearly periodic, with the length of one of the decaying oscillations being the pitch period, P (see figure 1.1b). The reciprocal of the pitch period is called the fundamental frequency or pitch. The frequency of the oscillation is approximately that of the major resonance of the vocal tract, while the bandwidth of the resonance determines the rate of decay of the oscillation [3]. The unvoiced segments (figure 1.1c) are those that seem to be random noise.

From the observation of the waveform, a reasonable model of the system would be that of figure 1.2. The time-varying linear system is excited by either a quasi-periodic train of impulses (of the proper fundamental frequency) for voiced sounds or random noise for unvoiced sounds. The digital filter represents the effect of the vocal tract, the glottal source, and the lips. The output of the filter is the speech waveform.

To represent the signal using the model, the form of the excitation signal and the parameters of the digital filter must be specified. Many reliable methods have been developed for the determination of the type of excitation function (impulses or random noise) and its characteristics (amplitude and pitch) [1,3,4,5]. The subject of this thesis is the specification and determination of the time-varying digital filter.

Usually the determination of the filter is simplified because

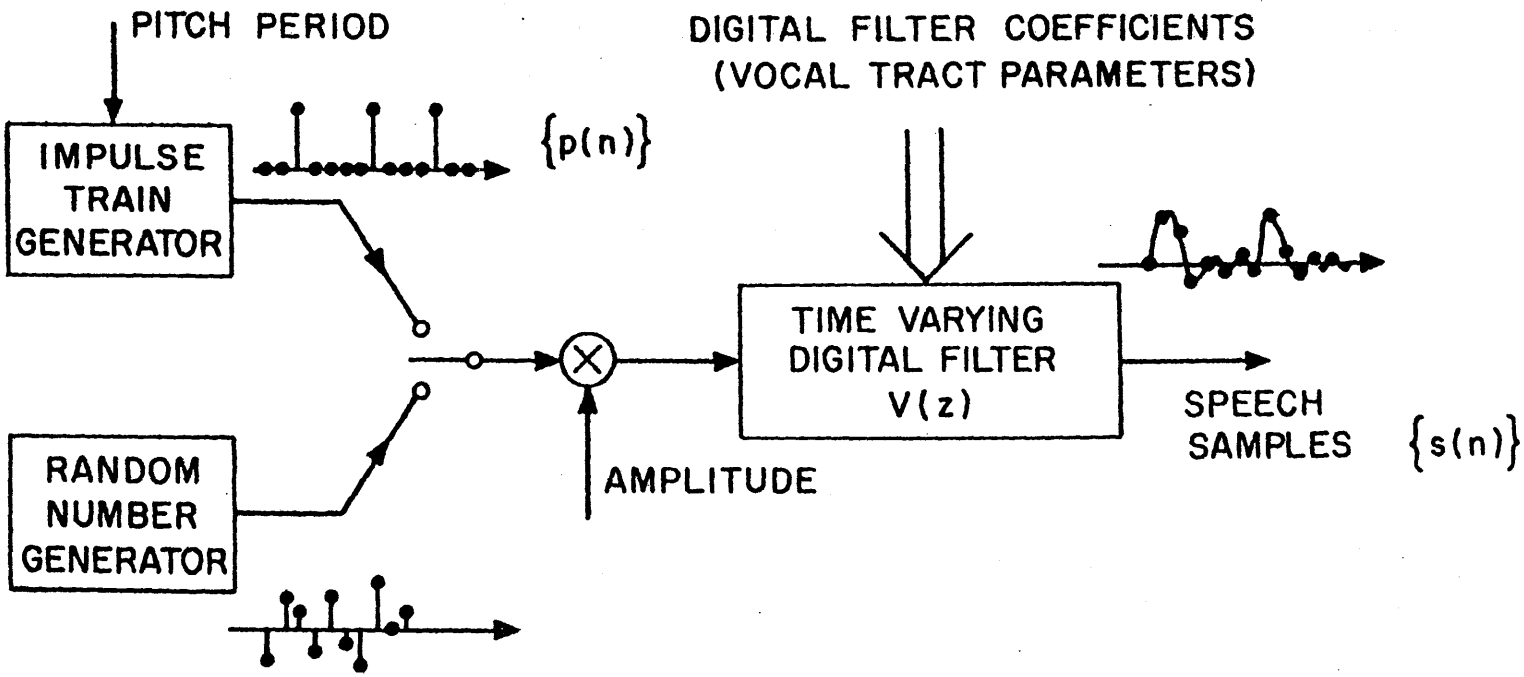

Figure 1.2 Speech Production Model
(From Kopec [2])

the speech signal is divided into short segments (10-30 msec) during which the signal is approximately stationary. For each one of these segments a time-invariant filter with constant coefficients can be used. Then the time-varying digital filter can be expressed as a filter with constant coefficients that are updated regularly throughout the speech signal.

The method of linear prediction has been used with much success to estimate the constant coefficients for the stationary segments of the waveform [3,6,7]. For linear predictive coding (LPC) the coefficients for a stationary segment are determined by minimizing a squared prediction error criterion. The LPC coefficients are easily obtained by solving a set of linear equations.

Since the assumption of stationarity is an approximation, a modification of LPC is examined in this thesis that enables the method to be used to analyze nonstationary signals. For the method (which we shall call time-varying LPC), each coefficient of the digital filter is allowed to change in time by assuming it is a linear combination of some set of known time functions. Using the same least squares error technique as used for LPC, the coefficients of the linear combinations of the time functions can be found by solving a set of linear equations. Therefore the determination of the digital filter parameters for time-varying LPC is similar to that for traditional LPC, but there is a larger number of coefficients that must be obtained for a given order model.

There are many possible advantages of time-varying LPC. The system model may be more realistic since it allows for the continuously changing behavior of the vocal tract. This should enable the model to have increased accuracy and sensitivity. In addition, the method may be more efficient since it will allow for the analysis of longer periods of speech. Therefore, even though time-varying LPC involves a larger number of coefficients than traditional LPC, it will divide the speech signal into fewer segments. This could result in a possible reduction of the total number of parameters needed to accurately model a segment of speech for time-varying LPC as compared with regular LPC.

An interesting problem in itself is the question of how exactly to measure and assess the performance of the time-varying LPC estimation method. One of the goals of this thesis is to explore methods for understanding the time-varying models and for evaluating their performance.

1.1 Thesis Outline

In Chapter II the method of traditional LPC is reviewed and the method of time-varying LPC is developed. Chapter III contains a discussion of the computations needed for time-varying LPC.

In Chapter IV, the general characteristics of time-varying linear prediction are determined by using the method to analyze several synthetic data test cases. Chapter V presents the results obtained by using the method on actual speech data. Chapter VI summarizes the experimental results, notes the limitations of the method, and examines future research possibilities.

CHAPTER II

TIME-VARYING LINEAR PREDICTION

The speech production model discussed in the introduction is a linear digital filter excited by an input pulse train. One representation for the digital filter would be a general rational transfer function of the form

$$H(z) = G \frac{1 + \sum_{i=1}^{r} b_i z^{-i}}{1 + \sum_{i=1}^{p} a_i z^{-i}} \tag{2.1}$$

The parameters that describe the model are the coefficients (a_i, b_i) of the denominator and numerator and the gain factor G. To specify the system for a speech segment, the model parameters would need to be estimated from the speech samples. For the general transfer function that contains zeros as given by 2.1, the estimation of the parameters involves the solution of a set of nonlinear equations [1].

A simpler model and estimation problem arises by assuming that the order of the numerator polynomial is zero, so that the model reduces to an all-pole filter. As shown by Markel and Gray [3], the transfer function of the speech production model can be represented as

$$H(z) = G(z)\ V(z)\ L(z) \tag{2.2}$$

where G(z) is the glottal shaping model, V(z) is the vocal tract model, and L(z) is the lip radiation model. The glottal shaping model is of the form

$$G(z) = \frac{1}{(1 - e^{-aT}z^{-1})^2} \tag{2.3}$$

where T is the time between speech samples. The lip radiation model is

$$L(z) = 1 - z^{-1} \tag{2.4}$$

The vocal tract is modelled as a cascade of a number of two-pole resonators. Each resonance is called a format with an associated center frequency F_i and bandwidth B_i. Then the vocal tract model is

$$V(z) = \frac{1}{\prod_{i=1}^{k} (1-2e^{-\pi B_i T} \cos(2\pi F_i T) z^{-1} + e^{-2\pi B_i T} z^{-2})} \tag{2.5}$$

For these models, the total speech production transfer function is

$$H(z) = \frac{(1 - z^{-1})}{(1-e^{-aT}z^{-1})^2 [\prod_{i=1}^{k} (1-2e^{-\pi B_i T} \cos(2\pi F_i T) z^{-1} + e^{-2\pi B_i T} z^{-2})]} \tag{2.6}$$

However since aT is usually much less than unity, the numerator term nearly cancels one of the glottal terms $(1 - e^{-aT}z^{-1})$ in the denominator. Then the model can be further simplified by assuming an all-pole synthesis model of the form

$$H(z) = \frac{G}{(1 + \sum_{i=1}^{p} a_i z^{-i})} \qquad (2.7)$$

The model is justified under many conditions, although nasal sounds may require a model with zeros as provided in equation 2.1 [3].

For the all-pole synthesis model, the speech signal s(n) at time n is given as a linear combination of the past p speech samples and the input u(n)

$$s(n) = -\sum_{i=1}^{p} a_i s(n-i) + G\, u(n) \qquad (2.8)$$

For a given speech signal $(s(n), n=0,1,\ldots,N-1)$ the coefficients a_i, the gain factor G, and the pitch period of the input u(n) for the model of 2.8 need to be determined.

The method of linear prediction (or linear predictive coding LPC) has been used to estimate the coefficients and the gain factor [3,6,7]. For LPC, it is assumed that the signal is stationary over the time interval of interest and therefore the coefficients given in the model of equation 2.8 are constants. This is a reasonable approximation over short intervals (10-30 msec).

Of course, the speech waveform never matches such a model exactly, and in particular, the assumption of piecewise stationarity is an obvious idealization. Since the vocal tract is changing throughout each utterance a more realistic model would be one that is time-varying. Therefore for the method of time-varying linear predictive coding that is presented in this thesis, the all-pole filter coefficients are allowed to change with time. Since there is a strong relationship between LPC and time-varying LPC, the method of estimating the filter coefficients by LPC will be reviewed first.

2.1. Linear Prediction

The use of linear prediction in speech processing is well documented in Markel and Gray [3] and Makhoul [6]. This section will follow the derivation given by Makhoul. Additional information may be found in the references given above.

For the model of the speech signal, the input sequence u(n) is completely unknown. However, it can be seen from equation 2.8 that a speech sample can be approximately predicted in terms of the past samples by

$$\hat{s}(n) = -\sum_{i=1}^{p} a_i \, s(n-i) \qquad (2.9)$$

The error sequence from the predictor is given by

$$e(n) = s(n) - \hat{s}(n)$$

$$= s(n) + \sum_{i=1}^{p} a_i \, s(n-i) \qquad (2.10)$$

where the terms a_i, $i=1,\ldots,p$, are the predictor coefficients.

The method of least squares estimation can be applied to estimating the predictor coefficients. By this method, the coefficients are obtained that minimize the total squared error

$$E = \sum_n e^2(n) = \sum_n \left(s(n) + \sum_{i=1}^{p} a_i \, s(n-i)\right)^2 \qquad (2.11)$$

where the limits of the summation over n are left unspecified for the moment. This optimization criterion is chosen because it results in easily solved linear equations and it gives excellent results for speech analysis [3].

The total error is minimized by setting the partial derivative with respect to each coefficient equal to zero

$$\frac{\partial E}{\partial a_j} = 2\sum_n \left(s(n) + \sum_{i=1}^{p} a_i \, s(n-i)\right) s(n-j) = 0 \qquad (2.12)$$

which reduces to

$$\sum_{i=1}^{p} a_i \sum_n s(n-i)s(n-j) = -\sum_n s(n)s(n-j) \qquad (2.13)$$

$$1 \leq j \leq p$$

By defining

$$c(i,j) = \sum_{n} s(n-i)s(n-j) \qquad (2.14)$$

the set of equations for the coefficients given by 2.13 becomes

$$\sum_{i=1}^{p} a_i\, c(i,j) = -c(0,j) \qquad 1 \leq j \leq p \qquad (2.15)$$

This set of p linear equations must be solved for the p predictor coefficients. These are two specific methods for the estimation of the parameters arising from different choices for the range of summation over n.

For the <u>covariance method</u>, it is assumed that there are N speech samples available (s(n), n=0,1,...,N-1). The first sample that can be predicted in terms of the past p samples is s(p). Therefore the error is minimized over the interval [p,N-1]. For the covariance method the coefficient c(i,j) is given by

$$c(i,j) = \sum_{n=p}^{N-1} s(n-i)s(n-j) \qquad \begin{array}{l} 0 \leq i \leq p \\ 1 \leq j \leq p \end{array} \qquad (2.16)$$

which is the covariance of the signal s(n). This is called the covariance method because the coefficients c(i,j) in 2.15 form a covariance matrix. From 2.16, it can be seen that the covariance coefficients are symmetric

$$c(i,j) = c(j,i) \tag{2.17}$$

The <u>autocorrelation method</u> assumes that the error is minimized over an infinite time interval. The coefficients of 2.14 become

$$\begin{aligned} c(i,j) &= \sum_{n=-\infty}^{\infty} s(n-i)s(n-j) \\ &= \sum_{n=-\infty}^{\infty} s(n)s(n+|i-j|) \\ &= r(|i-j|) \end{aligned} \tag{2.18}$$

The coefficients for the autocorrelation method are only a function of $|i-j|$. The set of equations given by 2.14 reduces to

$$\sum_{i=1}^{p} a_i\, r(|i-j|) = -r(j) \qquad i \le j \le p \tag{2.19}$$

Since the signal $s(n)$ is known only over a finite interval $[0,N-1]$, $s(n)$ is defined as being zero for $n < 0$, or $n \geq N$. Then $r(\ell)$ is given as

$$r(\ell) = r(-\ell) = r(|i-j|) = \sum_{n=0}^{N-1-|\ell|} s(n)s(n+|\ell|) \tag{2.20}$$

which is the definition for the short term autocorrelation for the delay $\ell = |i-j|$. Therefore this method is called the autocorrelation method.

Note that there are effectively discontinuities between the data inside and the data outside the interval [0,N-1] (since the signal s(n) is set equal to zero for $n < 0$ or $n \geq N$) and these discontinuities generally affect the determination of the coefficients. To show why this is so, we can compare the limits of the error summation for the autocorrelation method with the limits for the covariance method. It can be seen that the autocorrelation method attempts to predict more speech samples at each end of the interval than the covariance method does.

At the beginning of the interval, the autocorrelation method predicts

$$\hat{s}(n) = -\sum_{i=1}^{n} a_i\ s(n-i) \qquad 1 \leq n \leq p - 1 \qquad (2.21)$$

Since the predictor does not have p past speech values to use, the coefficients a_i will be distorted somewhat in order to reduce the predictor error for the first samples. Similarly, at the end of the interval, the method predicts

$$\hat{s}(n) = -\sum_{i=1}^{p} a_i\ s(n-i) \qquad N \leq n < N + p - 1 \qquad (2.22)$$

But since s(n) has been defined as zero for n=N, this causes distortion in the estimates of the coefficients because the system is attempting to predict an unrealistic signal.

Usually in order to reduce the effects due to the end discontinuities, the signal is multiplied by a window function w(n) (such as a Hamming window [7]) which goes to zero at both ends of the interval so that

$$s'(n) = w(n)s(n) \qquad 0 \leq n \leq N-1 \qquad (2.23)$$

$$= 0 \qquad \text{otherwise}$$

The window signal s'(n) is then used in equation 2.20 to define the autocorrelation coefficients. Markel and Gray [3] state that the speech signal should be windowed for either the covariance or the autocorrelation method when using data involving several pitch periods. The use of a window can reduce the spectral distortion caused by the end effects and may permit the estimation of more resonances in the spectrum. A more complete discussion concerning the use of windows for linear prediction is given in [3].

Both equations 2.15 and 2.19 are a set of p linear equations and p unknowns. They can be expressed in matrix form as

$$\Phi \, \underline{a} = -\underline{\Psi} \qquad (2.24)$$

For the covariance method the matrix Φ is symmetric and there is an efficient procedure called <u>Cholesky decomposition</u> for solving for the parameters [3]. For the autocorrelation method

$c(i,j) = c(i+1,j+1) = r(|i-j|)$ and the autocorrelation matrix form is

$$\begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(p-1) \\ r(1) & r(0) & r(1) & & \\ r(2) & r(1) & \cdot & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & r(1) \\ r(p-1) & & & r(1) & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = - \begin{bmatrix} r(1) \\ r(2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ r(p) \end{bmatrix} \qquad (2.25)$$

The matrix Φ is Toeplitz, for all the elements along any diagonal are equal. Because Φ is Toeplitz, there is an even more efficient method called <u>Levinson's recursion</u> for finding the predictor coefficients for the autocorrelation method [3].

The least squares method for determining the predictor coefficients is based on the assumption that the signal is deterministic. However other methods for estimating the parameters such as maximum likelihood or minimum variance estimation could be used by assuming the signal is a sample from a random process. These methods can be shown to yield the same solutions for the predictor coefficients as the least squares method [3,5].

2.2. Time-Varying Linear Prediction

For the method of time-varying linear prediction, the prediction coefficients are allowed to change with time, so that 2.8 becomes

$$s(n) = -\sum_{i=1}^{p} a_i(n)\, s(n-i) + Gu(n) \tag{2.26}$$

With this model, the speech signal is not assumed to be stationary and therefore the time-varying nature of the coefficient $a_i(n)$ must be specified.

The actual time variation of $a_i(n)$ is not known, however as suggested by Gelb [9], the coefficients can be approximated as a linear combination of some known functions of time, $u_k(n)$, so that

$$a_i(n) = \sum_{k=0}^{q} a_{ik}\, u_k(n) \tag{2.27}$$

With a model of this form the constant coefficients a_{ik} are to be estimated from the speech signal, where the subscript i is a reference to the time-varying coefficient $a_i(n)$, while the subscript k is a reference to the set of time functions $u_k(n)$. Without any loss of generality, it is assumed that $u_0(n) = 1$. Possible sets of functions that could be used include powers of time

$$u_k(n) = n^k \tag{2.28}$$

or trigonometric functions as in a Fourier series

$$u_k(n) = \cos\,(k\omega n) \qquad k \text{ even}$$
$$u_k(n) = \sin\,(k\omega n) \qquad k \text{ odd} \tag{2.29}$$

where ω is a constant dependent upon the length of the speech data. Liporace [10] seems to have been the first to have formulated the problem as in equation 2.27. His analysis used the power series of the form of 2.28 for the set of functions.

From equations 2.26 and 2.27, the predictor equation is given as

$$\hat{s}(n) = -\sum_{i=1}^{p} \left(\sum_{k=0}^{q} a_{ik}\, u_k(n) \right) s(n-i) \tag{2.30}$$

and the prediction error is

$$\begin{aligned} e(n) &= s(n) - \hat{s}(n) \\ &= s(n) + \sum_{i=1}^{p} \left(\sum_{k=0}^{q} a_{ik}\, u_k(n) \right) s(n-i) \end{aligned} \tag{2.31}$$

As in LPC, the criterion of optimality for the coefficients is the minimization of the total squared error

$$E = \sum_n e^2(n) = \sum_n \left(s(n) + \sum_{i=1}^{p} \sum_{k=0}^{q} a_{ik}\, u_k(n) s(n-i) \right)^2 \tag{2.32}$$

when the limits are again left unspecified.

The error is minimized with respect to each coefficient by setting

$$\frac{\partial E}{\partial a_{j\ell}} = 2\sum_n [s(n) + \sum_{i=1}^{p} \sum_{k=0}^{q} a_{ik}\, u_k(n)\, s(n-i)] u_\ell(n)\, s(n-j) = 0 \qquad \begin{matrix} 1 \le j \le p \\ 0 \le \ell \le q \end{matrix} \tag{2.33}$$

By rearranging 2.33 and changing the order of the summation, the equations for the coefficients become

$$\sum_{i=1}^{p} \sum_{k=0}^{q} a_{ik} [\sum_n u_k(n)\, u_\ell(n)\, s(n-i)\, s(n-j)] = -\sum_n u_\ell(n)\, s(n)\, s(n-j) \qquad \begin{matrix} 1 \le j \le p \\ 0 \le \ell \le q \end{matrix} \tag{2.34}$$

By defining

$$c_{k\ell}(i,j) = \sum_n u_k(n)\, u_\ell(n)\, s(n-i)\, s(n-j) \tag{2.35}$$

2.34 can be rewritten as

$$\sum_{i=1}^{p} \sum_{k=0}^{q} a_{ik}\, c_{k\ell}(i,j) = -c_{0\ell}(0,j) \qquad \begin{matrix} 1 \le j \le p \\ 0 \le \ell \le q \end{matrix} \tag{2.36}$$

For the coefficient $c_{k\ell}(i,j)$, the subscripts k and ℓ refer to the set of time functions, while the variables inside the parentheses, i and j, refer to the signal samples. Since $u_0(n) = 1$, the time-varying LPC coefficients $c_{00}(i,j)$ are the same as the LPC coefficients c(i,j) given by equation 2.16.

The minimization of the total error results in a p(q+1) set of equations that must be solved for the coefficients a_{ik}. The form of 2.36 is very similar to that of equation 2.8 for the LPC coefficients. The time-varying LPC equations reduce to the LPC equations for q=0, that is when $a_i(n)$ is a constant, $a_i(n) = a_{i0}$.

The limits of the sum over n can be chosen to correspond to the limits for the covariance and autocorrelation methods of LPC given earlier. For the covariance method, the sum over n goes from p to N-1 so that the elements of the matrix become

$$c_{k\ell}(i,j) = \sum_{n=p}^{N-1} u_k(n)\, u_\ell(n)\, s(n-i)\, s(n-j) \qquad (2.37)$$

For the covariance method, the following elements are equal

$$c_{k\ell}(i,j) = c_{\ell k}(i,j) = c_{k\ell}(j,i) = c_{\ell k}(j,i) \qquad (2.38)$$

Equation 2.36 can be expressed in matrix form by defining the vectors

$$\underline{a}_i^T = [a_{1i}, a_{2i}, a_{3i}, \ldots, a_{pi}] \qquad 0 \le i \le q \tag{2.39}$$

and

$$\underline{\psi}_i^T = [c_{0i}(0,1), c_{0i}(0,2), \ldots, c_{0i}(0,p)] \qquad 0 \le i \le q \tag{2.40}$$

and the matrix

$$\Phi_{k\ell} = \begin{bmatrix} c_{k\ell}(1,1) & c_{k\ell}(1,2) & \cdots & c_{k\ell}(1,p) \\ c_{k\ell}(2,1) & c_{k\ell}(2,2) & & \\ & & & \\ & & & \\ c_{k\ell}(p,1) & & & c_{k\ell}(p,p) \end{bmatrix} \tag{2.41}$$

$$0 \le k \le q$$
$$0 \le \ell \le q$$

From equations 2.38 and 2.41 it is clear that $\Phi_{k\ell} = \Phi_{\ell k} = \Phi_{k\ell}^T$ so that equation 2.36 becomes

$$\begin{bmatrix} \Phi_{00} & \Phi_{01} & \cdots & \Phi_{0q} \\ \Phi_{10} & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \Phi_{q0} & & & \Phi_{qq} \end{bmatrix} \begin{bmatrix} \underline{a}_0 \\ \underline{a}_1 \\ \underline{a}_2 \\ \\ \\ \underline{a}_q \end{bmatrix} = - \begin{bmatrix} \underline{\psi}_0 \\ \underline{\psi}_1 \\ \underline{\psi}_2 \\ \\ \\ \underline{\psi}_q \end{bmatrix} \qquad (2.42)$$

or

$$\Phi \underline{A} = -\underline{\Psi} \qquad (2.43)$$

This is a matrix equation that must be solved for the coefficient vector $\underline{A}$. Because $\Phi_{k\ell} = \Phi_{\ell k} = \Phi_{k\ell}^T$, the matrix Φ is block symmetric matrix with symmetric blocks. In this arrangement Φ is a (q+1)x(q+1) matrix composed of pxp blocks.

Equation 2.36 can also be expressed so that Φ is a (pxp) block symmetric matrix with (q+1)x(q+1) symmetric blocks by defining

$$\tilde{a}_i^T = [a_{i0}, a_{i1}, \ldots, a_{iq}] \qquad 1 \leq i \leq p \qquad (2.44)$$

$$\tilde{\psi}_i = [c_{00}(0,i), c_{01}(0,i), \ldots, c_{0q}(0,i)] \quad 1 \leq i \leq p \qquad (2.45)$$

and

$$\Phi(i,j) = \begin{bmatrix} c_{00}(i,j) & c_{01}(i,j) & c_{oq}(i,j) \\ c_{10}(i,j) & . & . \\ . & . & . \\ c_{q0}(i,j) & . & c_{qq}(i,j) \end{bmatrix} \qquad (2.46)$$

$$1 \leq i \leq p$$

$$1 \leq j \leq p$$

then 2.36 becomes

$$\begin{bmatrix} \Phi(1,1) & \Phi(1,2) & \cdots \\ \Phi(2,1) & & \\ . & & \\ . & & \\ \Phi(p,1) & & \Phi(p,p) \end{bmatrix} \begin{bmatrix} \underline{\tilde{a}}_1 \\ \underline{\tilde{a}}_2 \\ . \\ . \\ \underline{\tilde{a}}_p \end{bmatrix} = - \begin{bmatrix} \underline{\tilde{\psi}}_1 \\ \underline{\tilde{\psi}}_2 \\ . \\ . \\ \underline{\tilde{\psi}}_p \end{bmatrix} \qquad (2.47)$$

or

$$\Phi\underline{\tilde{A}} = -\underline{\tilde{\Psi}} \qquad (2.48)$$

To develop a method similar to the autocorrelation method, the error must be minimized over an infinite time interval. The equation for the coefficients is

$$c_{k\ell}(i,j) = \sum_{n=-\infty}^{\infty} u_k(n)\, u_\ell(n)\, s(n-i)\, s(n-j) \qquad (2.49)$$

Letting n' = n - i, this becomes

$$c_{k\ell}(i,j) = \sum_{n'=-\infty}^{\infty} u_k(n'+i)\ u_\ell(n'+i)\ s(n')\ s(n'+i-j) \qquad (2.50)$$

However by this definition $c_{k\ell}(i,j)$ is not a function of (i-j) alone. So the matrix formed by $c_{k\ell}(i,j)$ could not be expressed as a block Toeplitz matrix. By a slight change of definition the problem can be corrected. The time variation of the coefficients of 2.25 will be changed so that

$$a_i(n) = \sum_{k=0}^{q} a_{ik}\ u_k(n-i) \qquad 1 \le i \le p \qquad (2.51)$$

As an example of this, for the power series

$$a_i(n) = \sum_{k=0}^{q} a_{ik}(n-i)^k \qquad 1 \le i \le p \qquad (2.52)$$

where (n-i) is set to zero for i > n. By performing the minimization of 2.31 again the resulting equations are

$$\sum_{i=1}^{p} \sum_{k=0}^{q} a_{ik} \sum_{n} u_k(n-i)\ u_\ell(n-j)\ s(n-i)\ s(n-j) = -\sum_{n} u_\ell(n-j)s(n)s(n-j)$$

$$1 \le j \le p \qquad (2.53)$$

$$0 \le \ell \le q$$

The autocorrelation coefficients can be defined as

$$c_{k\ell}(i,j) = \sum_{n=-\infty}^{\infty} u_k(n-i)\ u_\ell(n-j)\ s(n-i)\ s(n-j) \tag{2.54}$$

$$= \sum_{n=-\infty}^{\infty} u_k(n)\ u_\ell(n+i-j)\ s(n)\ s(n+i-j)$$

$$\triangleq r_{k\ell}(i-j)$$

The autocorrelation coefficients are cross-symmetric (a term used by Flinn [12] to express the symmetry relationships of $\gamma_{k\ell}$ and $\gamma_{\ell k}$) because using 2.53, we have

$$r_{k\ell}(m) = r_{k\ell}(i-j) = c_{k\ell}(i,j) \tag{2.55}$$

$$= \sum_{n=-\infty}^{\infty} u_k(n)\ u_\ell(n+m)\ s(n)\ s(n+m)$$

and with n' = n + m, this becomes

$$r_{k\ell}(m) = \sum_{n'=-\infty}^{\infty} u_k(n'-m) u_\ell(n') s(n') s(n'-m)$$

so

$$r_{k\ell}(m) = r_{\ell k}(-m) \tag{2.56}$$

but for $k \neq \ell$, $r_{k\ell}(m) \neq r_{\ell k}(m)$. With the definition of $r_{k\ell}(i-j)$, equation 2.53 is given as

$$\sum_{i=1}^{p} \sum_{k=0}^{q} a_{ik}\, r_{k\ell}(i-j) = -r_{0\ell}(-j) \qquad \begin{array}{l} 1 \le j \le p \\ 0 \le \ell \le q \end{array} \qquad (2.57)$$

which can be changed into matrix form by using 2.56 so that

$$\sum_{i=1}^{p} \sum_{k=0}^{q} r_{\ell k}(j-i)\, a_{ik} = -r_{0\ell}(-j) \qquad \begin{array}{l} 1 \le j \le p \\ 0 \le \ell \le q \end{array} \qquad (2.58)$$

By defining the following vectors

$$\underline{a}_i^T = [a_{1i}, a_{2i}, a_{3i}, \ldots, a_{pi}] \qquad 0 \le i \le q \qquad (2.59)$$

$$\underline{\psi}_i^T = [r_{0i}(-1), r_{0i}(-2), \ldots, r_{0i}(-p)] \qquad 0 \le i \le q \qquad (2.60)$$

and matrix

$$\Phi_{\ell k} = \begin{bmatrix} r_{\ell k}(0) & r_{\ell k}(-1) & r_{\ell k}(-2) & \cdots & r_{\ell k}(-p+1) \\ r_{\ell k}(1) & r_{\ell k}(0) & & & \\ r_{\ell k}(2) & r_{\ell k}(1) & & & \\ & & & & \\ r_{\ell k}(p-1) & & & & r_{\ell k}(0) \end{bmatrix}$$

$$0 \le \ell \le q \qquad (2.61)$$
$$0 \le k \le q$$

then equation 2.58 is

$$\begin{bmatrix} \Phi_{00} & \Phi_{01} & \Phi_{02} & \cdots & \Phi_{0q} \\ \Phi_{10} & \Phi_{11} & & & \\ \cdot & & & & \\ \cdot & & & & \\ \Phi_{q0} & & & & \Phi_{qq} \end{bmatrix} \begin{bmatrix} \underline{a}_0 \\ \underline{a}_1 \\ \underline{a}_2 \\ \\ \underline{a}_q \end{bmatrix} = - \begin{bmatrix} \underline{\psi}_0 \\ \underline{\psi}_2 \\ \underline{\psi}_3 \\ \\ \underline{\psi}_q \end{bmatrix} \qquad (2.62)$$

or

$$\Phi\underline{A} = -\underline{\Psi} \qquad (2.63)$$

For this problem, Φ is a (q+1)x(q+1) block matrix with each block ($\Phi_{\ell k}$) being Toeplitz (see equation 2.61). In addition $\Phi_{\ell k} = \Phi_{k\ell}^T$, because the autocorrelation coefficients are cross-symmetric as shown in equation 2.56. Equation 2.58 can also be arranged as a pxp block matrix with the blocks being (q+1)x(q+1) by defining

$$\underline{\tilde{a}}_i^T = [a_{i0}, a_{i1}, \ldots, a_{iq}] \qquad 1 \le i \le p \qquad (2.64)$$

$$\underline{\tilde{\psi}}_i^T = [r_{00}(-i), r_{01}(-i), \ldots, r_{0q}(-i)] \qquad 1 \le i \le p \qquad (2.65)$$

and

$$R(m) = \begin{bmatrix} r_{00}(m) & r_{01}(m) & \cdots & r_{0q}(m) \\ r_{10}(m) & & & \\ & & & \\ r_{q0}(m) & & & r_{qq}(m) \end{bmatrix} \quad -p \leq m \leq p \qquad (2.66)$$

Equation 2.58 becomes

$$\begin{bmatrix} R(0) & R(-1) & R(-2) & \cdots & R(-p+1) \\ R(1) & R(0) & & & \\ R(2) & R(1) & & & \\ & & & & \\ R(p-1) & & & & R(0) \end{bmatrix} \begin{bmatrix} \underline{\tilde{a}}_1 \\ \underline{\tilde{a}}_2 \\ \\ \\ \underline{\tilde{a}}_p \end{bmatrix} = - \begin{bmatrix} \underline{\tilde{\psi}}_1 \\ \underline{\tilde{\psi}}_2 \\ \\ \\ \underline{\tilde{\psi}}_p \end{bmatrix} \qquad (2.67)$$

or

$$\underline{R}\underline{\tilde{A}} = -\underline{\tilde{\psi}} \qquad (2.68)$$

The matrix R is symmetric and block Toeplitz, but it is not block symmetric. From equation 2.66 it can be seen that $R(m) = R(-m)^T$.

Since the signal is only known between [0,N-1], it is assumed to be zero outside of this interval and equation 2.54 becomes

$$r_{k\ell}(i-j) = \sum_{n=-\infty}^{\infty} u_k(n)\, u_\ell(n+i-j)\, s(n)\, s(n+i-j) \qquad (2.69)$$

$$= \sum_{n=0}^{N-1-(i-j)} u_k(n)\, u_\ell(n+i-j)\, s(n)\, s(n+i-j) \quad i \geq j$$

The autocorrelation method can be viewed as a multichannel filtering problem as considered by Wiggins and Robinson [11] and Flinn [12]. With this interpretation the multichannel input is $s(n)$, $u_1(n)s(n)$, $u_2(n)s(n)$, ..., $u_q(n)s(n)$, and the output is the one-step predicted estimate of $s(n)$.

The covariance and autocorrelation methods of time-varying LPC have been given these names because they have the same range for the summation of the squared error as the corresponding methods in traditional LPC. However the same physical interpretations of the elements $c_{k\ell}(i,j)$ and $r_{k\ell}(i-j)$ as given in LPC cannot be used for time-varying LPC. The element $c_{k\ell}(i,j)$ could be interpreted as the covariance of the signals $u_k(n)s(n)$ and $u_\ell(n)s(n)$. However since the signal is not assumed to be stationary, it is not possible to give a similarly meaningful "autocorrelation" interpretation for the autocorrelation elements.

The limits of the error minimization for the time-varying covariance method have been chosen so that the squared error is summed only over those speech samples that can be predicted from the past p samples. However, the error for the time-varying autocorrelation method is minimized over the entire time interval (the same range that is used for the traditional LPC autocorrelation method). Therefore, the same discussion concerning the distortions of the LPC coefficients due to the discontinuities in the data at the ends of the interval apply to the time-varying coefficients.

This distortion in the coefficients estimated by the autocorrelation method may or may not be significant depending on the data at the ends of the interval.

It was noted that the windowing of the speech signal is a usual practice for the LPC correlation method in order to reduce the distortion. However even though windowing might reduce the end effects for the autocorrelation method, it also imposes an additional time variation upon the speech sample. This can cause two problems. The estimates of the coefficients by time-varying LPC will be adversely affected since the method by its very formulation is sensitive to any time variation of the system parameters such as that caused by the windowing of the signal. In addition, the window affects the relative weight of the errors throughout the interval. Since the windowed data at both ends of the interval will be smaller, there is more signal energy in the central data. Therefore the minimization of the error will result in coefficients that in general will reproduce the signal in the center of the interval better than at the ends.

Because of distortion in the estimates caused by the end effects when the data is not windowed and the possible adverse effects on the estimates when the data is windowed, the autocorrelation method seems to have more disadvantages than the covariance method. Since a window will have the same distortive effect for the covariance method, the use of a window does not seem beneficial.

This will be discussed in more detail in Chapter IV.

The method developed in this chapter estimates only the coefficients of the time-varying filter of equations 2.26 and 2.27. The method does not give an estimate of the gain factor, G, of equation 2.26, (which for this method should be time-varying); however, the regular LPC method can estimate the gain factor based on the minimized errors [3,6]. The effects of not having a time-varying gain on the resulting analysis are shown and discussed in Chapter V.

In closing, it should be noted that the error summation method used by Liporace [10], does not correspond exactly to either of the two methods discussed in this chapter. In his method, the error is minimized over all the data in the interval, however he does not modify the definition of the time-varying coefficients in order to create "autocorrelation" coefficients. In addition, he does not discuss whether the data outside the interval should be set to zero, or whether a window should be used for his method.

CHAPTER III

COMPUTATIONAL ASPECTS OF TIME-VARYING LINEAR PREDICTION

For the time-varying linear prediction method outlined in Chapter 2, the predictor coefficients (a_{ik}, $i \leq i \leq p$, $0 \leq k \leq q$) are obtained by solving a set of linear equations, given by equation 2.36 which is repeated here

$$\sum_{i=1}^{p} \sum_{k=0}^{q} a_{ik}\, c_{k\ell}(\ell,j) = -c_{0\ell}(0,j) \qquad 1 \leq j \leq p, \quad 0 \leq \ell \leq q \tag{3.1}$$

This can be expressed in matrix form as (see equation 2.43)

$$\Phi \underline{A} = -\underline{\Psi} \tag{3.2}$$

where $\underline{A}$ is a vector of the coefficients.

Because the number of coefficients increases linearly with the number of terms in the series expansion (q+1), the increase in the amount of computation for time-varying LPC as compared with LPC (where q=0) is significant. This chapter will discuss the computational aspects of time-varying linear prediction.

The computations necessary for the determination of the coefficients can be divided into two categories. Much of the computational effort is involved with calculating the elements

$c_{k\ell}(i,j)$ for Φ and $\underline{\Psi}$. The rest of the operations are needed for taking the inverse of the $p(q+1)$ square matrix Φ to obtain the predictor coefficient vector $\underline{A}$. Each category will be examined separately.

3.1. Computation of the Matrix Coefficients

There are $p^2(q+1)^2$ elements in the matrix Φ and $p(q+1)$ elements in the vector $\underline{\Psi}$. However Φ is symmetric for both the covariance and the autocorrelation methods that were discussed in Chapter 2. Therefore, the largest number of matrix elements that need to be calculated is $\frac{p(q+1)[p(q+1)+1]}{2}$. But because Φ may have additional symmetry, this number can be reduced further. In addition, the computational burden can be reduced because some elements of the matrix can be calculated easily from other elements that have been previously determined.

For the covariance method, the matrix elements are given by (eq. 2.37)

$$c_{k\ell}(i,j) = \sum_{n=p}^{N-1} u_k(n)\, u_\ell(n)\, s(n-i)\, s(n-j) \qquad (3.3)$$

with

$$c_{k\ell}(i,j) = c_{\ell k}(i,j) = c_{k\ell}(j,i) = c_{\ell k}(j,i)$$

As it was noted in Chapter 2, the set of linear equations can be expressed as a block symmetric matrix equation with each block being a symmetric matrix. (The matrix Φ can either be expressed as a (q+1)x(q+1) block matrix with pxp blocks or as a pxp block matrix with (q+1)x(q+1) blocks.) Because of this symmetry only $\frac{p(p+1)}{2} \frac{(q+1)(q+2)}{2}$ elements for the matrix Φ need to be calculated. Also, many of the elements can be calculated from previously computed elements without having to sum over all the data as given in equation 3.3. For example, for $k=\ell=0$ it is easy to see that [3]

$$c_{00}(i,j) = \sum_{n=p}^{N-1} s(n-i)\, s(n-j) \tag{3.4}$$

$$= c_{00}(i-1,j-1) + s(p-i)\, s(p-j) - s(n-i)\, s(n-j)$$

With this recursion only the coefficients $c_{00}(0,j)$, $0 \leq j \leq p$ require the complete summation of equation 3.3. The rest of the coefficients $c_{00}(i,j)$, $1 \leq i, j \leq p$ can be calculated using equation 3.4.

Recursions can also be developed for the elements when $k \neq 0$ or $\ell \neq 0$. For example, for the power series expansion where $u_r(n) = n^r$, the matrix elements are

$$c_{k\ell}(i,j) = \sum_{n=p}^{N-1} n^{k+\ell}\, s(n-i)\, s(n-j) \tag{3.5}$$

As an example, when $k + \ell = 1$

$$c_{10}(i,j) = c_{01}(i,j) = \sum_{n=p}^{N-1} n\ s(n-i)\ s(n-j) \tag{3.6}$$

Letting n' = n-1

$$c_{10}(i,j) = \sum_{n'=p-1}^{N-2} (n'+1)\ s(n'+1-i)\ s(n'+1-j) \tag{3.7}$$

$$= \sum_{n'=p-1}^{N-2} n's(n'+1-i)\ s(n'+1-j) + \sum_{n'=p-1}^{N-2} s(n'+1-i)\ s(n'+1-j)$$

But the last two terms can be seen to be given as

$$\sum_{n'=p-1}^{N-2} n's(n'-i+1)s(n'-j+1) = c_{10}(i-1,j-1) + (p-1)s(p-i)s(p-j) - (N-1)s(N-i)s(N-j) \tag{3.8}$$

and

$$\sum_{n'=p-1}^{N-2} s(n'-i+1)s(n'-j+1) = c_{00}(i-1,j-1) + s(p-i)s(p-j) - s(N-i)s(N-j)$$

By using 3.8, equation 3.7 becomes

$$c_{10}(i,j) = c_{10}(i-1,j-1) + c_{00}(i-1,j-1) + p\ s(p-i)s(p-j) - Ns(N-i)s(n-j) \tag{3.9}$$

which gives a simple recursion for $c_{10}(i,j)$. In general for $k+\ell=m$,

$$c_{m0}(i,j) = \sum_{n=p}^{N-1} n^m\ s(n-i)\ s(n-j) \tag{3.10}$$

$$= \sum_{n'=p-1}^{N-2} (n'+1)^m \, s(n-i+1)\, s(n-j+1)$$

$$= \sum_{n'=p}^{N-1} (n'+1)^m \, s(n-i+1)s(n-j+1) + p^m s(p-i)s(p-j)$$

$$- N^m s(N-i)s(N-j)$$

By using the binomial expression

$$(n+1)^m = \sum_{r=0}^{m} \binom{m}{r} n^{m-r} \qquad (3.11)$$

where

$$\binom{m}{r} = \frac{m!}{(m-r)!r!}$$

we obtain

$$c_{m0}(i,j) = \sum_{n=p}^{N-1} \left(\sum_{r=0}^{m} \binom{m}{r} n^{m-r} \right) s(n-i+1)s(n-j+1) \qquad (3.12)$$

$$+ p^m s(p-i)s(p-j) - N^m s(N-i)s(N-j)$$

$$= \sum_{r=0}^{m} \binom{m}{r} c_{m-r,0}(i-1,j-1) + p^m s(p-i)s(p-j)$$

$$- N^m s(N-i)s(N-j)$$

which gives the recursion for $c_{m0}(i,j)$.

The power series covariance method has the additional advantage that for $k+\ell=m$

$$c_{m0}(i,j) = c_{k\ell}(i,j) \tag{3.13}$$

so that only the elements $c_{k0}(i,j)$, $0 \leq k \leq 2q$, need to be computed.

It should be noted that for the power series case, the matrix Φ of equation 2.40 is a (q+1)x(q+1) block Hankel matrix (where all the (pxp) matrices along the secondary diagonal, northeast to southwest, are equal). This is significant when attempting to invert Φ efficiently to obtain the predictor coefficient vector, $\underline{A}$. This will be discussed in more detail later in the chapter.

Table 3.1 summarizes the reduction of computations for the covariance power series method as well as several others yet to be discussed. Column 2 lists the indices of $c_{k\ell}(i,j)$ that must be calculated for the matrix Φ and the vector $\underline{\Psi}$. Column 3 lists the only elements that need to be calculated by summing over all the data as given in equation 3.5. The number of elements that can be calculated in terms of the elements previously computed elements are given in column 4. The rest of the elements of the matrix can be found by using the symmetry equations. The computation of the remaining elements involves just a few more operations. For the determination of each one of the elements listed in column 3, the summation involves approximately N additions and (k+1)N

TABLE 3.1

MATRIX COMPUTATIONAL EFFORT FOR TIME-VARYING LPC

Method	Indices of Elements to be Calculated	Indices of Elements to be Calculated by Summation [total number]	Number of Elements to be Determined Recursively
Covariance* Power Series $c_{k\ell}(i,j)$	$1 \le i,j \le p \quad 0 \le k,\ell \le q$ $i=0 \quad j=0 \quad k=0 \quad 0 \le \ell \le q$	$i=0 \quad 1 \le j \le p \quad 0 \le k \le q \quad \ell=0$ $i=1 \quad j=1 \quad 0 \le k \le q \quad \ell=0$ $i=1 \quad 1 \le j \le p \quad q \le k \le 2q \quad \ell=0$ $[p(2q+1)+q+1]$	$qp^2 + \frac{p(p+1)}{2} - q - 1$
Covariance Fourier Series $c_{k\ell}(i,j)$	as above	$i=0 \quad 1 \le j \le p \quad 0 \le k \le q \quad \ell=0$ $i=1 \quad j=1 \quad 0 \le k \le q \quad \ell=0$ $i=1 \quad 1 \le j \le p \quad 1 \le k \le q \quad 1 \le \ell \le k$ $\left[p\left(\frac{q^2+3q+2}{2}\right)+q+1\right]$	$\left(\frac{q^2+3q+2}{2}\right)\left(\frac{p^2-p}{2}\right)+(q+1)(p-1)$
Autocorrelation* (for either series) $r_{k\ell}(m)$	$-p+1 \le m \le p-1 \quad 0 \le k,\ell \le p$ $m=-p \quad k=0 \quad 0 \le \ell \le q$	$0 \le m \le p-1 \quad 0 \le k \le q \quad 0 \le \ell \le k$ $m=p \quad 0 \le k \le q \quad \ell=0$ $\left[p\left(\frac{q^2+3q+2}{2}\right)+q+1\right]$	$\frac{q(q+1)}{2}(p-1)$

*The computational effort for the corresponding LPC method can be found by using q=0.

multiplications, where k is the index denoting the power of n used in the summation.

The elements for the covariance method with the Fourier series expansion can also be calculated recursively. The Fourier series time functions are

$$u_r(n) = \cos(r\omega n) \qquad r \text{ even} \tag{3.14}$$
$$= \sin(r\omega n) \qquad r \text{ odd} \qquad 0 \le r \le q$$

The constant ω can be chosen to be $\frac{2\pi}{N}$ or $\frac{\pi}{N}$, where N is the total number of speech data points. If $\omega = \frac{2\pi}{N}$, the time-varying coefficients will be the same at each end of the interval. However for $\omega = \frac{\pi}{N}$, this constraint is eliminated. A discussion of the differences between these constants will be given in Chapter 4.

To show the type of recursion for the Fourier series, the element for k=1, ℓ=0 is

$$c_{10}(i,j) = \sum_{n=p}^{N-1} \sin\left(\frac{\pi}{N} n\right) s(n-i)s(n-j) \tag{3.15}$$

$$= \sum_{n'=p-1}^{N-2} \sin\left(\frac{\pi}{N}(n'+1)\right) s(n'+1-i)s(n'+1-j)$$

$$= \sum_{n'=p}^{N-1} \sin\left(\frac{\pi}{N}(n'+1)\right) s(n'+1-i)s(n'+1-j)$$

$$+ \sin\left(\frac{\pi}{N} p\right) s(p-i)s(p-j) - \sin(\pi)\, s(N-i)s(N-j)$$

and by expanding the sine term

$$c_{10}(i,j) = \cos\frac{\pi}{N}\sum_{n'=p}^{N-1} \sin\left(\frac{\pi}{N}n'\right) s(n'+1-i)s(n'+1-j) \qquad (3.16)$$

$$+ \sin\frac{\pi}{N}\sum_{n'=p}^{N-1} \cos\left(\frac{\pi}{N}n'\right) s(n'+1-i)s(n'+1-j)$$

$$+ \sin\left(\frac{\pi}{N}p\right) s(p-i)s(p-j)$$

so

$$c_{10}(i,j) = \cos\frac{\pi}{N} c_{10}(i-1,j-1) + \sin\frac{\pi}{N} c_{20}(i-1,j-1) \qquad (3.17)$$

$$+ \sin\left(\frac{\pi}{N}p\right) \sin\,(p-i)\ s(p-j)$$

Similarly $c_{20}(i,j)$ can be found in terms of $c_{10}(i-1,j-1)$ and $c_{20}(i-1,j-1)$. Recursions for larger values of k and ℓ can be found, although the form of the recursions cannot be expressed as compactly as for the covariance power recursion of equation 3.13. It is also easy to see that the symmetry equation 3.14 for the power method elements is not true for the Fourier method. Therefore more elements must be calculated for the Fourier covariance matrix than for the power covariance matrix, as shown in column 3 of table 3.1. The summation for the covariance Fourier elements of column 3 involves approximately N additions, 3N multiplications and 2N trigonometric evaluations (for $k\geq 1$ and $\ell\geq 1$). There are N fewer multiplications and N fewer trigonometric evaluations for the elements with either k=0 or ℓ=0.

The autocorrelation method has matrix elements that are given by (see equation 2.66)

$$r_{k\ell}(m) = c_{k\ell}(i,j) \sum_{n=0}^{N-1-m} u_k(n)u_\ell(n+m)s(n)s(n+m) \qquad m = (i-j) \geq 0 \qquad (3.18)$$

with $r_{k\ell}(m) = r_{\ell k}(-m)$. Because the elements are only a function of i-j, a smaller number of elements need to be calculated by equation 3.18.

The elements for the autocorrelation method can also be calculated recursively in order to save computations. For the power series method, $u_r(n) = n^r$, and

$$r_{k\ell}(m) = \sum_{n=0}^{N-1-m} n^k(n+m)^\ell s(n)s(n+m) \qquad m \geq 0 \qquad (3.19)$$

With k=1, ℓ=0, this becomes

$$r_{10}(m) = r_{01}(-m) = \sum_{n=0}^{N-1-m} n\, s(n)s(n+m) \qquad (3.20)$$

However for k=0, ℓ=1,

$$r_{01}(m) = r_{10}(-m) = \sum_{n=0}^{N-1-m} (n+m)s(n)s(n+m) \qquad (3.21)$$

$$= \sum_{n=0}^{N-1-m} n\, s(n)s(n+m) + m \sum_{n=0}^{N-1-m} s(n)s(n+m)$$

so that

$$r_{01}(m) = r_{10}(m) + m\, r_{00}(m) \tag{3.22}$$

This illustrates the type of recursion for the power autocorrelation method elements. A general form for the recursions can be found by using equation 3.19 and the formula for the binomial expansion.

As an example of the recursion for the Fourier series, for $k=1$, $\ell=0$

$$r_{10}(m) = \sum_{n=0}^{N-1-m} \sin\left(\frac{\pi}{N}n\right) s(n)\; s(n+m) \tag{3.23}$$

and with $k=0$, $\ell=1$

$$r_{01}(m) = \sum_{n=0}^{N-1-m} \sin \frac{\pi}{N}(n+m)\; s(n)\; s(n+m) \tag{3.24}$$

$$= \cos\left(\frac{\pi}{N}m\right) \sum_{n=0}^{N-1-m} \sin\left(\frac{\pi}{N}n\right) s(n)\; s(n+m)$$

$$+ \sin\left(\frac{\pi}{N}m\right) \sum_{n=0}^{N-1-m} \cos\left(\frac{\pi}{N}n\right) s(n)\; s(n+m)$$

$$= \cos\left(\frac{\pi}{N}m\right) r_{10}(m) + \sin\left(\frac{\pi}{N}m\right) r_{20}(m)$$

General recursion formulas can be found for other values of k and ℓ so that an element with $\ell>k$ can be expressed in terms of the elements with $k\geq\ell$.

The number of elements that must be calculated for the autocorrelation methods are shown in Table 3.1. The summation for each autocorrelation power or Fourier element takes approximately the same number of operations as for the corresponding covariance power or Fourier element.

From the table it can be seen that $q>0$, the power covariance will take the least amount of computations for determining the matrix elements because of its special symmetry given by equation 3.13. The autocorrelation methods result in slightly more calculations and the Fourier covariance method needs the most computations.

Since the computation of a trigonometric function is more complex than the evaluation of an integer raised to a power, each method (covariance or autocorrelation) using the Fourier series will take longer than the same method using the power series.

There is another advantage of the power series method for the situation when the time-varying coefficients for an interval of speech data have been estimated and the interval is to be increased to include new data. The new matrix elements for the power series method can be calculated by using the matrix elements that were computed for the smaller interval and adding on the appropriate sums of the new data. However for the Fourier series methods, the period of the coefficients is dependent upon the interval of the data.

The addition of more data changes the interval length and the constant ω. The new matrix elements must be calculated by summation over all the data using the new ω. There is no way to use the matrix elements that were computed for the smaller interval (except for the elements with $k=\ell=0$, which are not dependent on ω). Of course, if the data is being windowed the matrix elements for the power series method also have to be totally recalculated.

3.2. Solution of the Equations

The solution of the equations is simplified due to the symmetry of the matrix. All of the methods so far discussed result in symmetric matrices. Therefore Cholesky decomposition can be used to invert the matrices to obtain the predictor coefficients. For a $(q+1)p\times(q+1)p$ matrix this will take $\frac{1}{3}(q+1)^3p^3+2(q+1)^2p^2+\frac{8}{3}(q+1)p-2$ operations [3]. Since the number of computations increase approximately as $(q+1)^3$, for very large q the computational burden is significantly greater than for traditional LPC using the covariance method where $q=0$. In addition, the constant LPC autocorrelation method for $q=0$ can use Levinsons recursion to solve the matrix equation. This method needs $p^2 - \frac{3}{2}p$ computations [3], so at least at first glance, it appears that the time-varying LPC method increases the number of computations by approximately $p(q+1)^3$ as compared with the constant LPC autocorrelation method.

However there are ways to exploit the symmetrical form of the matrix equation in order to further reduce the computations. For

For the autocorrelation methods, the matrix Φ can be arranged as a pxp block Toeplitz matrix with (q+1)x(q+1) matrices as elements. To solve this set of equations, a method which is an extension of Levinson's recursion algorithm to the multichannel filtering problem can be used [11]. This method is a special case of Rissanen's algorithm for the decomposition of block Toeplitz matrices. The multichannel Levinson's recursion requires $O((q+1)^3p^2)$ operations.

From the discussion in this chapter, it can be seen that generally more computations are needed for determining the elements of the matrices than for solving the equations. For example with p=10, q=2, N=1000, the number of computations needed to set up the matrix for the covariance power method is well over 100,000, while the number of computations used for solving the equations by Cholesky decomposition (which is the least efficient method) is less than 12,000. For this same case, the Fourier series method will be less efficient than the power series because of the additional time it takes to compute the trigonometric functions.

In general, it seems that time-varying LPC would involve more computations to accurately represent a given segment of nonstationary speech than would be needed for regular LPC, for which the speech segment has been divided into quasi-stationary intervals. Whether this increase is excessively large is not known.

CHAPTER IV

EXPERIMENTAL RESULTS FOR SYNTHETIC DATA

For the evaluation of time-varying linear prediction, the method was used to analyze synthetic data created by all-pole filters with known time-varying coefficients. The purpose of these test cases was to determine the general characteristics of time-varying LPC and to obtain some insight into methods for evaluating the performance of time-varying parameter identification techniques.

The first set of test cases was generated by all-pole filters with each coefficient changing as a truncated power or Fourier series. Therefore for these cases, the form of the system model of the time-varying linear prediction analysis matched the actual system generating the data. The results of these cases indicated the differences between using the power or Fourier series for analysis, between using the covariance or autocorrelation method of error summation (as developed in Chapter II), and between windowing or not windowing the signal.

The signal shown in figure 4.1 was generated by a 6 pole filter (p=6) with each time-varying coefficient being a quadratic power series (q=2). We shall call this a 6-2 power series filter. For example, a 6-0 filter is one with 6 poles and constant coefficients such as one used for regular LPC, and a 6-4 power series filter is one where the highest power in the series for each coefficient is n^4. A 6-2 Fourier series filter has one constant term, one sine term and

Figure 4.1 Synthetic Speech Example Generated by 6-2 Power Series Filter

one cosine term in the series for each coefficient.

The sampling rate for this (and for all the synthetic examples of this chapter) was 10 KHz. The "pitch period" of the excitation impulse train was 100 samples, corresponding to a fundamental frequency of 100 Hz. The signal length was 2000 samples, corresponding to a time interval of .2 sec.

For the evaluation of time-varying linear prediction using the different options, the "trajectories of the time-varying poles" of the all-pole filters were compared. By time-varying poles, we mean the zeros of p(z,n) (for each n in the interval [0,N-1]), where p(z,n) is defined as (from equation 2.26)

$$p(z,n) = 1 + \sum_{i=1}^{p} a_i(n)\, z^{-i} \qquad (4.1)$$

Note that in the time-varying case, the time-varying poles do not have the same significance as poles for a time-invariant filter. However when these "poles" change slowly in time, one should be able to deduce some qualitative aspects of the system behavior by observing the "pole trajectories". Hence we have used the ability of our parameter estimation system to track these poles as one possible measure of performance.

Using this comparison method does not imply that two filters with different pole trajectories are necessarily significantly different in impulse response or general characteristics. Instead, the comparison of the pole trajectories of the filters using the coefficients estimated by time-varying LPC with the pole trajectories of the filter generating the data will show qualitatively the effect of the different options on the accuracy of the analysis. The poles of the filters for each instant of time were calculated by Muller's method [19].

Figures 4.2 and 4.3 show the pole trajectories of the filters using the estimated coefficients. The graphs plot the real part of each pole on the ordinate and the imaginary part on the abscissa. The location of each pole of the filter is plotted every 25 msec of the analysis interval. The unit circle is also shown on the graphs for comparison purposes.

The angle of each pole, θ, is related to the center frequency, F, of the corresponding formant in the vocal tract model given in Chapter 2 by $F = \theta/2\pi T$ where T is time between samples. The radius of each pole, r, is related to the formant bandwidth, B, by $B = -(\ln r)/\pi T$.

Figures 4.2a shows the pole trajectories for the 6-2 filter estimated by using the covariance power series method with no windowing. Since these trajectories matched the pole trajectories of the generating filter so well, the original trajectories are not shown. Figure 4.2b shows the trajectories for the estimated 6-2 covariance power series filter using a Hamming window. The two trajectories 4.2a and 4.2b are only slightly different, illustrating the small effect of windowing for this example. The main differences

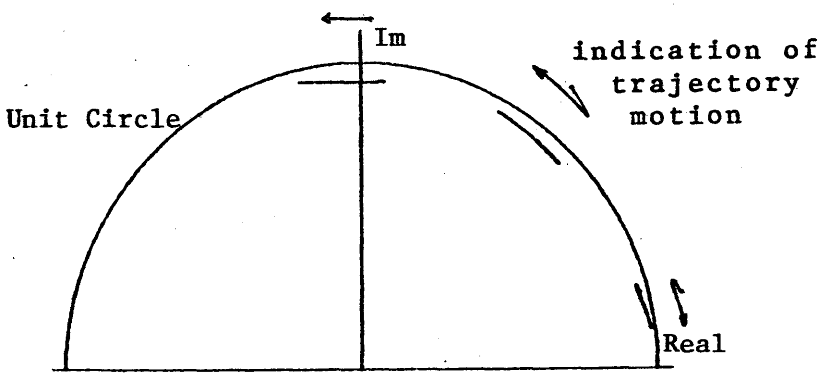

Figure 4.2a 6-2 Covariance Power Filter
(without window)

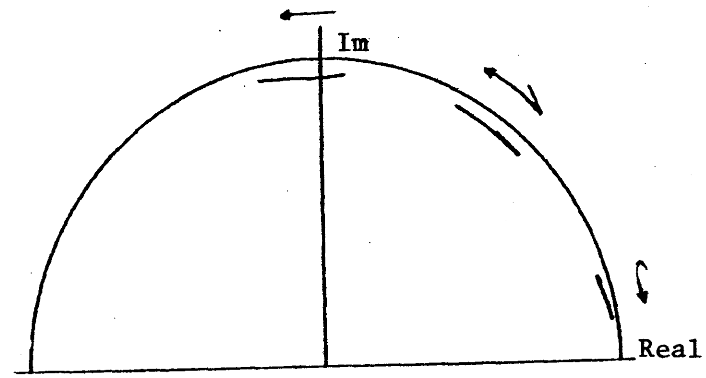

Figure 4.2b 6-2 Covariance Power Filter
(with window)

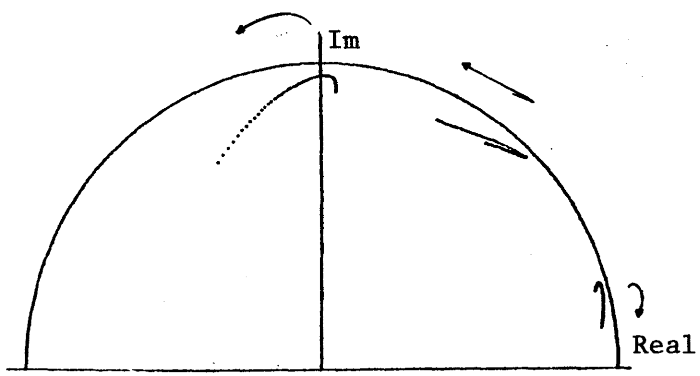

Figure 4.2c 6-2 Autocorrelation Power Filter
(without window)

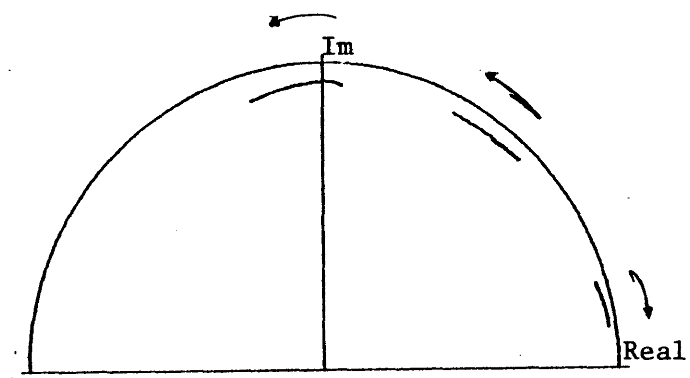

Figure 4.2d 6-2 Autocorrelation Power Filter
(with window)

Figure 4.2 Pole Trajectories for Power Series occur at each end of the trajectory, where the effect of the window is the most significant.

Figures 4.2c and 4.2d are the pole trajectories for the filters estimated by the 6-2 autocorrelation power series method with no windowing and windowing respectively. The general characteristics of the trajectories for the autocorrelation method without windowing are correct, but there is also a considerable amount of trajectory distortion. This is most evident in the third pole (the poles are numbered by having the one with the smallest angle be the first, etc.) where both the angle and radius of the pole at the end of the interval differ significantly from the correct values as shown in figure 4.2a. This would seem to verify the discussion at the end of Chapter II, where it was said that since the autocorrelation method attempted to minimize (unrealistically) the error at the extreme ends of the interval, there might be some distortion in the coefficients at the ends.

Figure 4.2d shows the pole trajectories for the filter for the 6-2 autocorrelation power series method with windowing. The windowing reduces the effect of the errors at the ends of the interval and therefore the pole trajectories are not as distorted as for those of figure 4.2c. In fact, these trajectories compare favorably with those of figures 4.2a and 4.2b. The only major differences are those of the third pole.

Figure 4.3 shows the trajectories for the filters estimated by the time-varying method using a 6-2 Fourier series (with $\omega = \frac{\pi}{N}$).

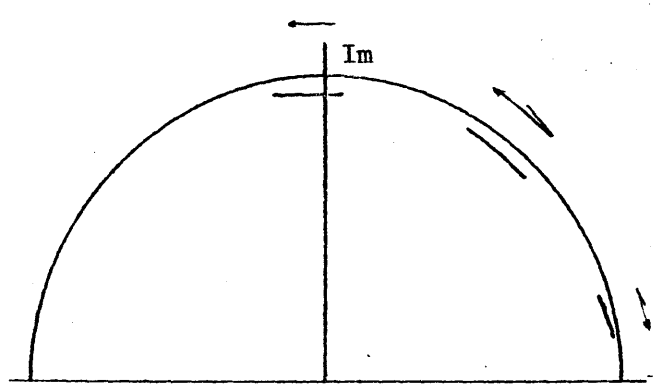

Figure 4.3a   6-2 Covariance Fourier Filter (without window)

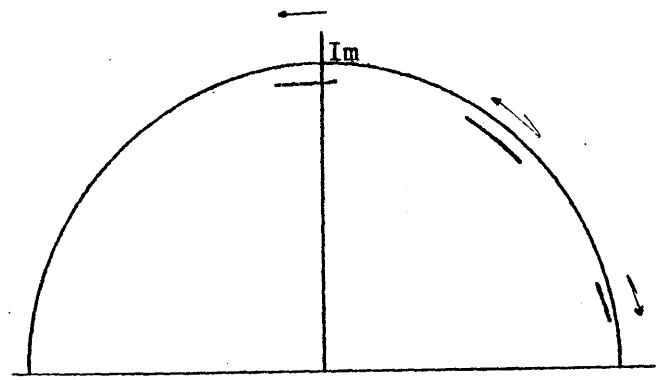

Figure 4.3b   6-2 Covariance Fourier Filter (with window)

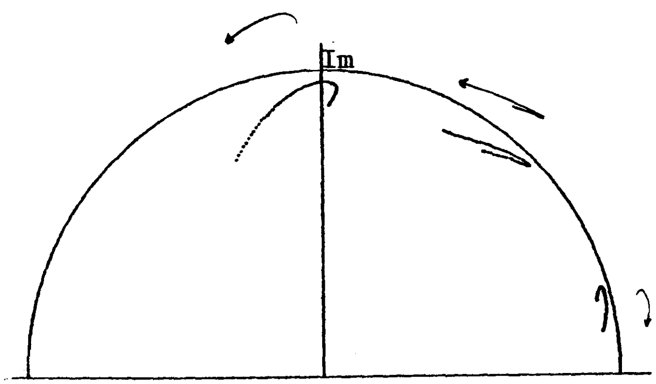

Figure 4.3c   6-2 Autocorrelation Fourier Filter (without window)

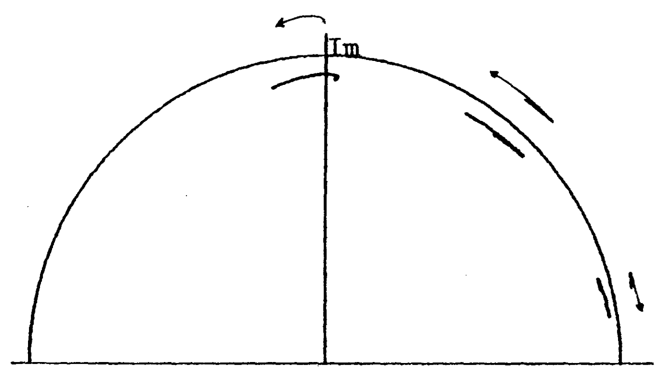

Figure 4.3d   6-2 Autocorrelation Fourier Filter (with window)

Figure 4.3 Pole Trajectories for Fourier Series

Each plot is remarkably similar to the corresponding plot of figure 4.2 for the power series method. For the 6-2 Fourier covariance method, both the non-windowed method shown in figure 4.3a and the windowed method shown in figure 4.3b differ significantly only for the third pole.

The 6-2 Fourier autocorrelation method without windowing (figure 4.3c) yields poles that show the same type of distortion as for the 6-2 power autocorrelation method. The use of a window for the Fourier autocorrelation method (figure 4.3d) again reduces the distortion.

To illustrate how well the pole trajectories of the Fourier method can match those of the original trajectories (generated by a power series filter), the pole angles (or formant center frequencies) for both trajectories are shown in figure 4.4. Figure 4.4a shows the center frequencies of the three poles for the estimated 6-2 covariance Fourier method without windowing as compared with the poles of the 6-2 power filter used to generate the data. The only significant differences between the two occur at the ends of the interval. By using the 6-4 covariance Fourier method even these slight differences can be removed. The center frequency trajectories for the estimated 6-4 covariance Fourier (shown in figure 4.4b) are nearly identical with the original trajectories.

The Fourier analysis methods shown so far have used a constant ω of $\frac{\pi}{N}$. However, in Chapter II, it was noted that a constant ω of

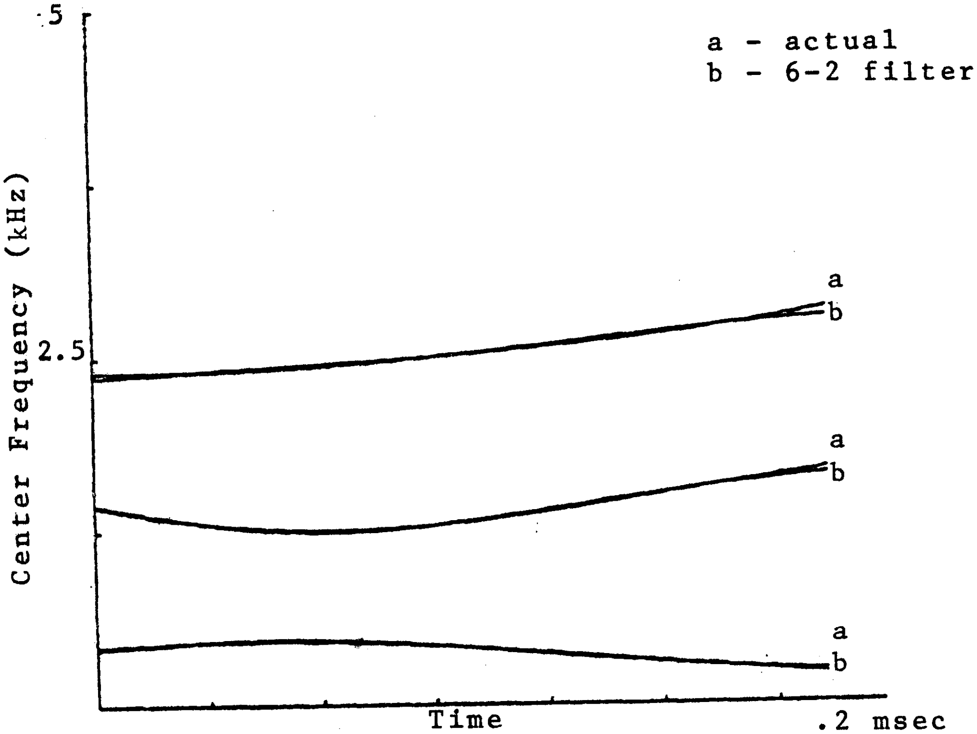

Figure 4.4a Center Frequency Trajectories
6-2 Covariance Fourier Filter
($\omega = \pi/N$)

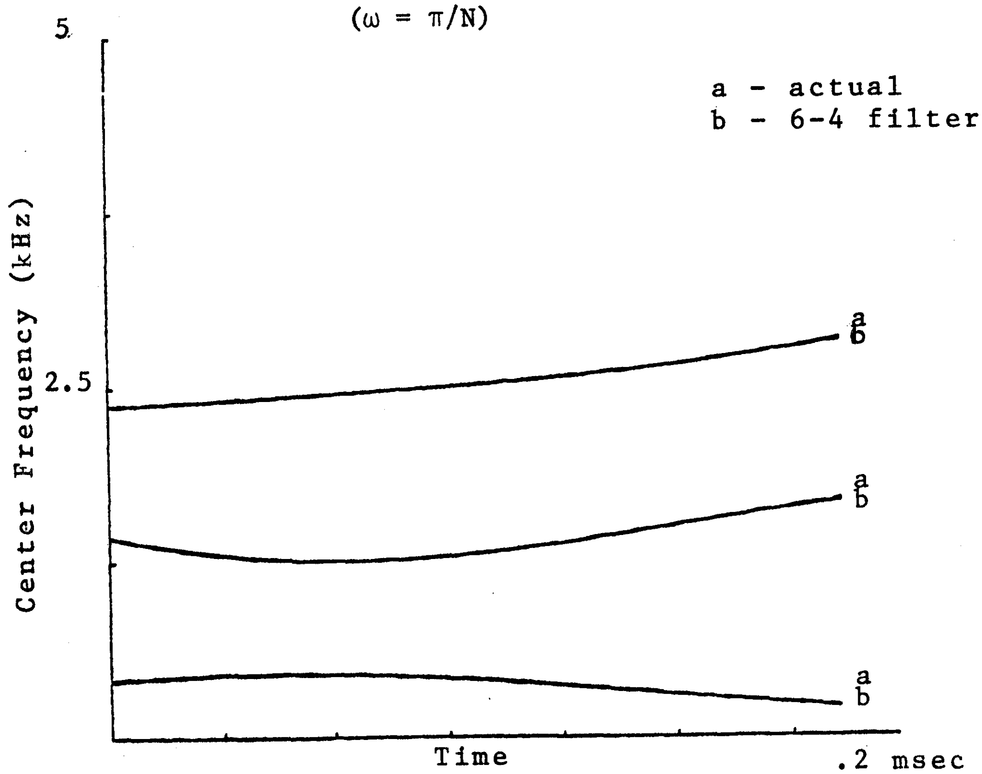

Figure 4.4b Center Frequency Trajectories
6-4 Covariance Fourier Filter
($\omega = \pi/N$)

$\frac{2\pi}{N}$ could also be used, but that the coefficients would be constrained to be the same at both ends of the interval. To illustrate the effect of using this constant, the pole trajectories for the analysis of the data of figure 4.1 estimated by a 6-4 covariance Fourier filter without windowing and with $\omega = \frac{2\pi}{N}$ are shown in figure 4.5a. It is easy to see that there are significant differences as compared with the trajectories of figure 4.2a. The three center frequency trajectories for the 6-4 Fourier method and the original 6-2 power generating filter are shown in figure 4.5b. There are differences in the estimated poles throughout the interval with significant distortion at both ends because the 6-4 Fourier filter is constrained to have the same poles at the ends of the interval. The center frequencies of the estimated 6-4 Fourier filter with $\omega = \frac{\pi}{N}$ as shown in figure 4.4b are clearly more accurate than the center frequencies of 6-4 Fourier filter with $\omega = \frac{2\pi}{N}$.

To demonstrate further the differences between the different options the pole center frequency trajectories for all the methods are shown in figure 4.6. Figure 4.6a shows the pole center frequencies for the methods which didn't window the signal. The major differences in the center frequencies occur at the ends with significant deviations for the autocorrelation methods. Figure 4.6b is a plot of the center frequencies for the methods using a Hamming window on the data. The use of the window tends to reduce the differences between the methods.

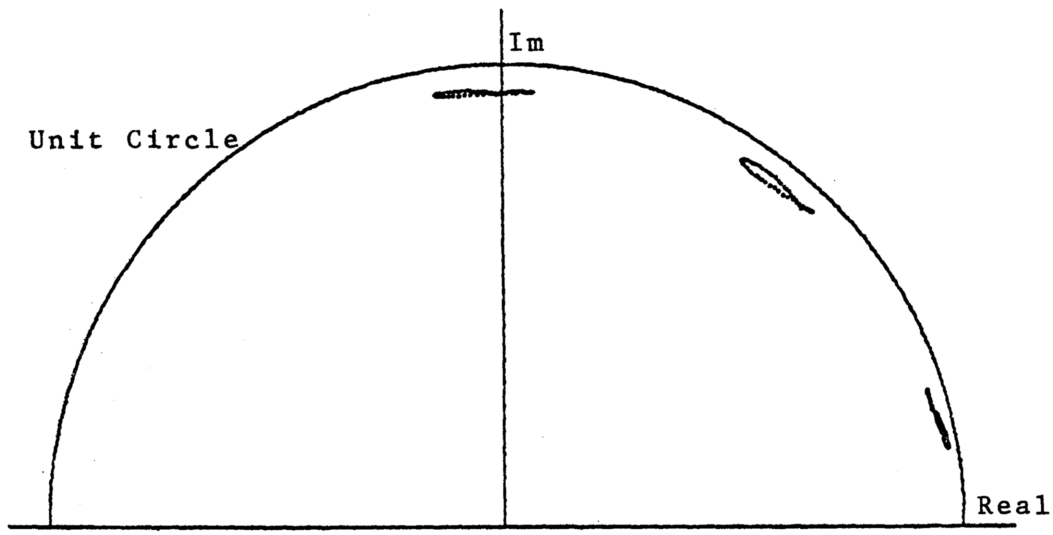

Figure 4.5a Pole Trajectories for
6-4 Covariance Fourier Filter
($\omega = 2\pi/N$)

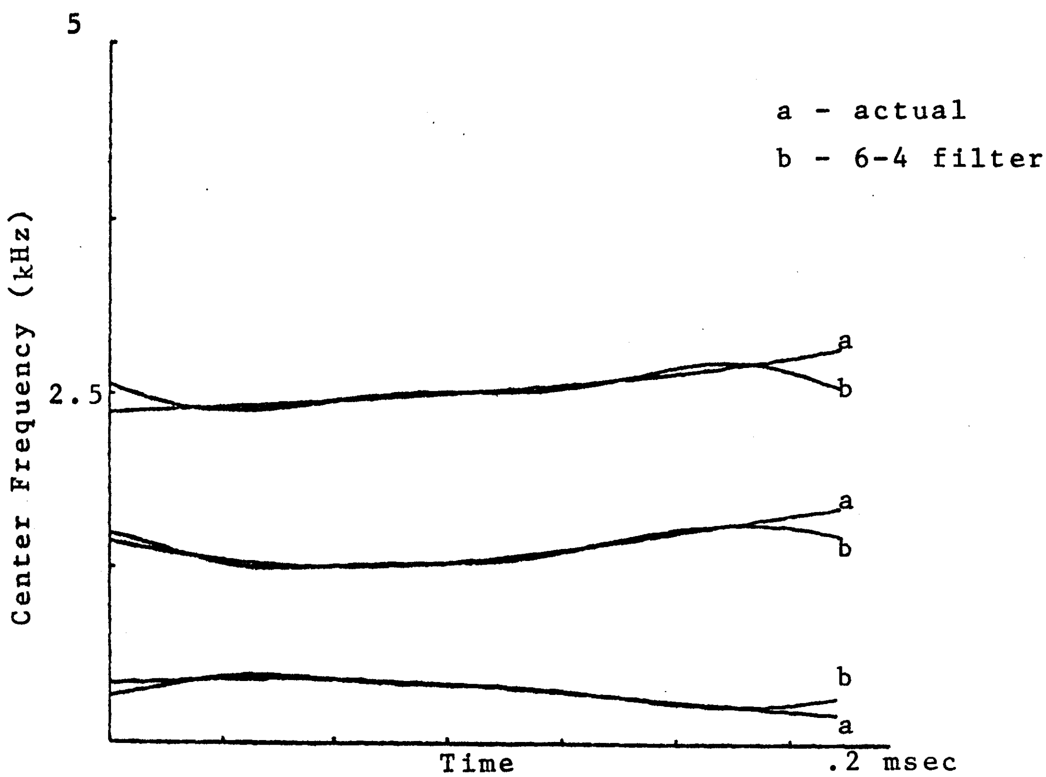

Figure 4.5b Center Frequency Trajectories
6-4 Covariance Fourier Filter
($\omega = 2\pi/N$)

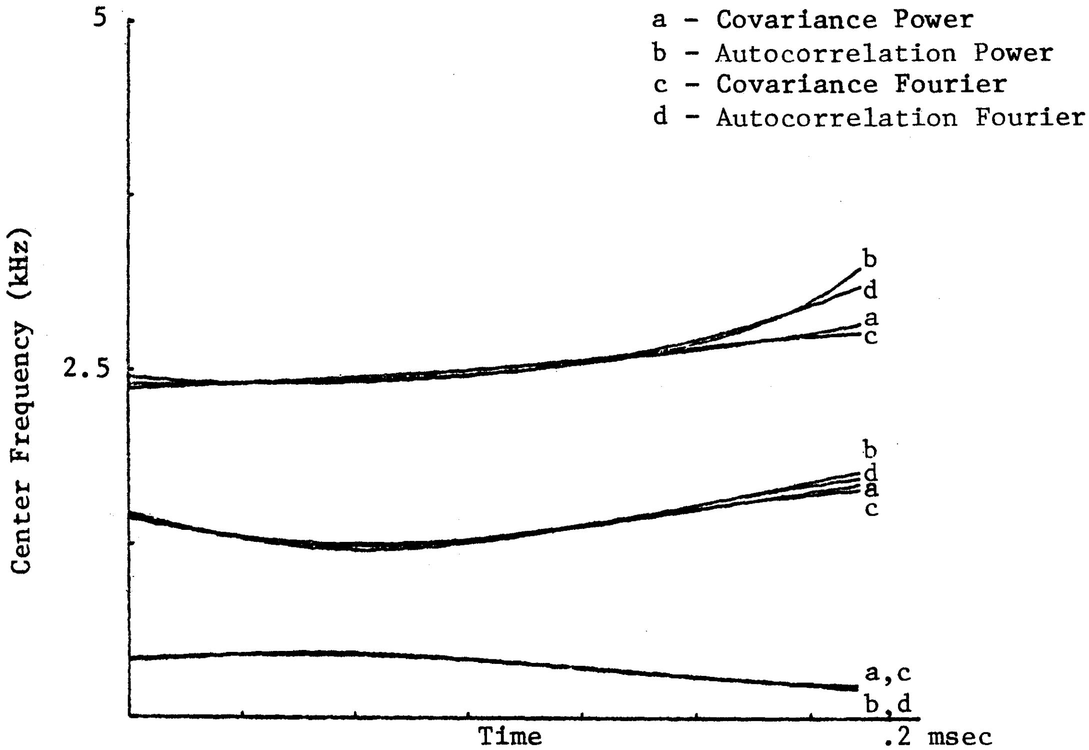

Figure 4.6a Center Frequency Trajectory
(6-2 Power Data, Not Windowed)

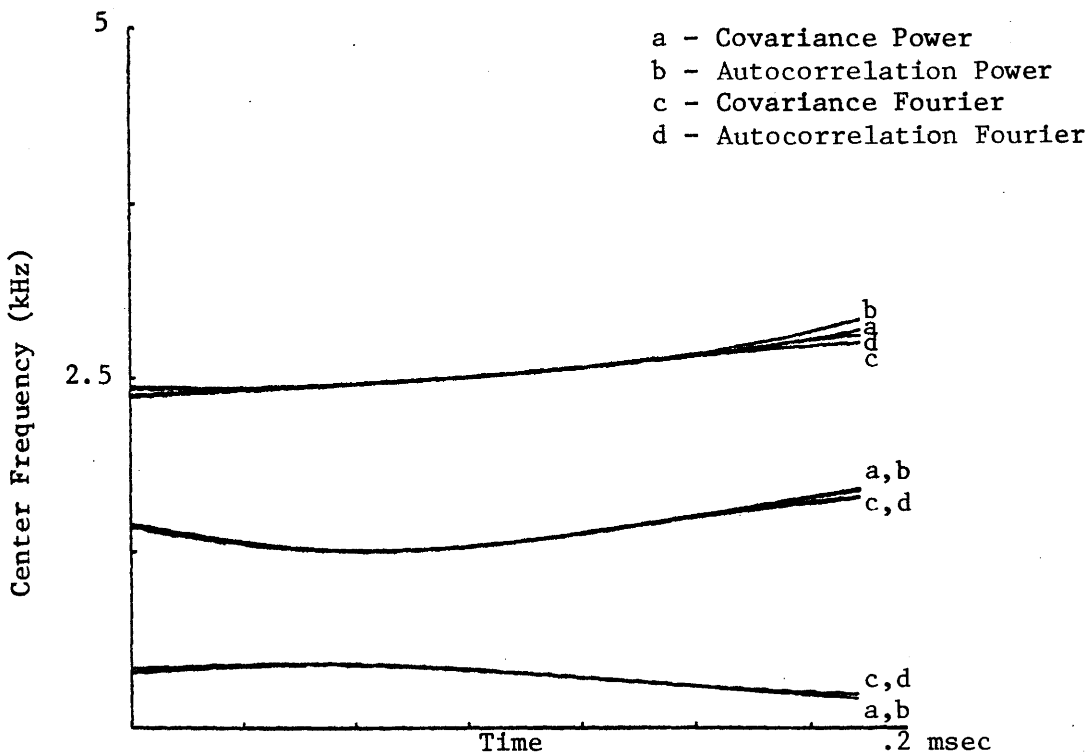

Figure 4.6b Center Frequency Trajectory
(6-2 Power Data, Windowed)

As we said earlier, the differences in the pole trajectories are not necessarily significant. Therefore the impulse train used to generate the data was passed through some of the estimated filters. The response of the estimated 6-2 covariance power filter (no windowing) was virtually identical to the original and therefore is not shown. The response of the 6-2 covariance Fourier filter (no windowing) is shown in figure 4.7a. There are very few differences between the 6-2 covariance Fourier filter response and the original data shown in figure 4.1. The 6-2 autocorrelation power filter response (no windowing) is shown in figure 4.7b and the 6-2 autocorrelation power filter response (windowing) is shown in figure 4.7c. It can be seen that the major differences between the responses of the autocorrelation filters and the original data occur at both ends of the interval. The autocorrelation response estimated without windowing the data does not match the original data as well as the autocorrelation response estimated with windowing, as we would expect from the pole trajectories of figures 4.2 and 4.3.

A similar test case was generated with a 6-2 Fourier filter ($\omega = \frac{\pi}{N}$) and the sample data is shown in figure 4.8. Time-varying linear prediction gave such similar pole trajectories for the different methods that the trajectories are not shown. Instead the pole center frequency trajectories for the estimation methods without windowing are shown in figure 4.9a and the pole center frequency trajectories for the methods with windowing are shown in

Figure 4.7a Response of 6-2 Covariance Fourier Filter
(without windowing the original data)

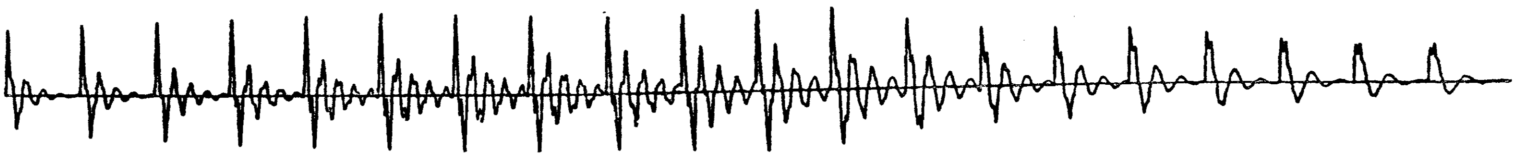

Figure 4.7b Response of 6-2 Autocorrelation Power Filter
(without windowing the original data)

Figure 4.7c Reponse of 6-2 Autocorrelation Power Filter
(with windowing the original data)

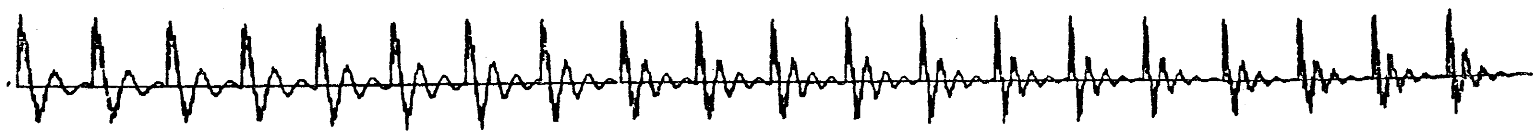

Figure 4.8 Synthetic Speech Example Generated by 6-2 Fourier Series Filter

figure 4.9b. The differences between the 6-2 Fourier and 6-2 power series and between the covariance and autocorrelation methods can be seen to be minor. In addition, the use of a window has only a small effect.

For this example the window has caused slightly more variation in the poles at the ends of the interval. However for the power series example shown in figure 4.6, the window reduced the end variation of the poles. This difference cannot be fully explained, but it does indicate that the general effects of windowing cannot be characterized precisely. Instead, the influence of the window on the resulting estimation is dependent to a large degree on the data in the interval and particularly to the data at each end.

There are many conclusions to be drawn from these examples. The differences between using a power series or a Fourier series for the analysis seems to be insignificant. In general, a filter using one series can be represented almost exactly by a filter using the other series with either the same or a slightly larger number of terms in the series. For example, the 6-2 power series filter could be represented accurately as a 6-4 Fourier series filter and a 6-2 Fourier series filter needed a 6-3 power series filter to represent it almost exactly.

The covariance method of summation gave better results than the autocorrelation method. Under some circumstances the differences between the two methods were minor, however this is not a general rule.

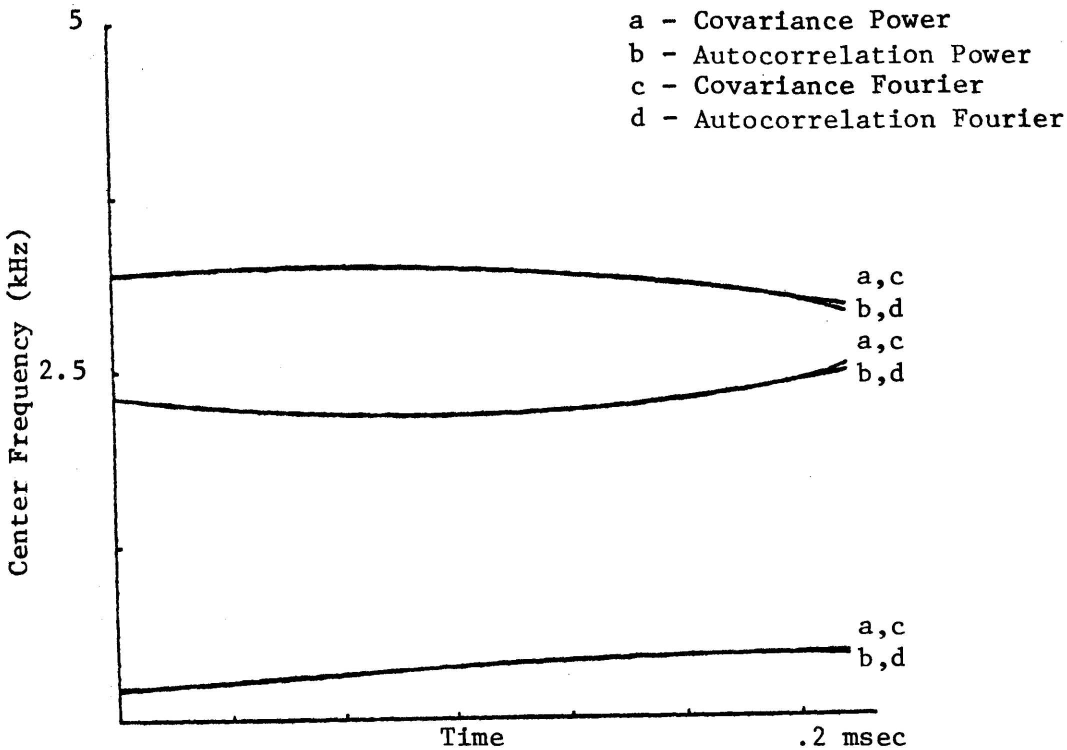

Figure 4.9a Center Frequency Trajectories (6-2 Fourier Data, Not Windowed)

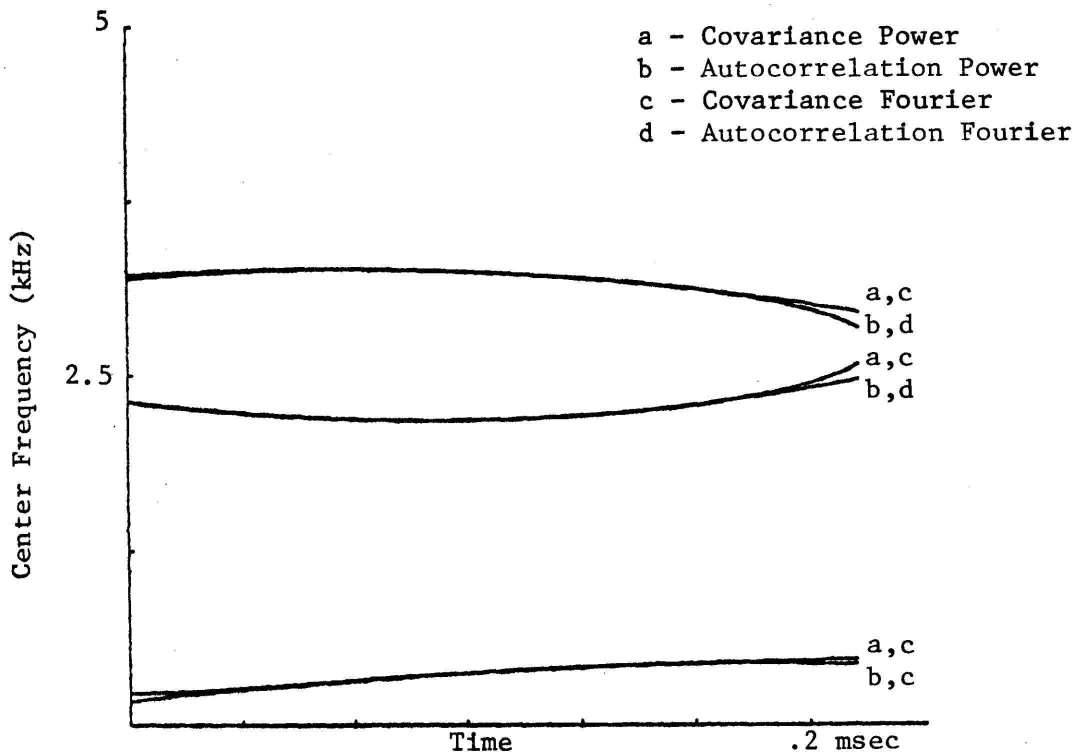

Figure 4.9b Center Frequency Trajectories (6-2 Fourier Data, Windowed)

The use of a window had only a slight affect on the analysis results. Windowing did not significantly degrade the performance of the covariance methods and in fact the autocorrelation methods that used a window seemed to give more accurate results than the autocorrelation methods without a window.

However, these results can be explained by the fact that the test cases were generated by a system whose form was the same as that of the analysis model. Therefore, these methods can estimate the coefficients of the series for the time-varying filter even with a window superimposed upon the signal because of the sample data in the central part of the interval.

However, actual speech signals are not generated by the system model of time-varying LPC and the use of a window will degrade the method's ability to track the time variation of the parameters accurately throughout the entire time interval. The basic problems with the use of a window were discussed in Chapter II, and, because of these problems, it does not seem that windowing is generally a good practice. In Chapter V, the effect of windowing actual nonstationary speech on the analysis results will be shown.

All of this analysis indicates that the covariance method without windowing should be used. Since the results seems to be similar for either the power or Fourier series, the power series is preferred because of its computational advantages over the Fourier series method as discussed in Chapter III.

The next set of cases involve the response of the system to step changes in the center frequency of the formants. These cases were generated by a four pole system. The center frequency of two poles changed discontinuously sometime during the interval.

The first case has one set of poles with a center frequency of 475 Hz and a bandwidth of 75 Hz, with the other set of poles having a center frequency of 1175 Hz and a bandwidth of 150 Hz. The sampling frequency was 10 KHz and the "pitch period" was 100 Hz. The length of the data was 600 samples (60 msec). At 30 msec, the center frequency of the 475 Hz poles was increased by a value ranging from 50 to 250 Hz. An example of the data for one test case is shown in figure 4.10 for the jump of 150 Hz (from 475 Hz to 625 Hz). The 6-3 covariance power method without windowing was used to analyze the data. Of interest is the trajectory of the center frequency of the first pole. The pole angle trajectories for different changes in the center frequencies are shown in figure 4.11a. The trajectory response for the time-varying linear prediction method is somewhat like the response of a low pass filter. However the response is anticipative since the entire interval is used to estimate the coefficients. In general the system response is almost homogeneous in that the pole angle trajectory for a given center frequency charge has a response that is approximately twice that of a pole trajectory for half the given frequency change.

The pole trajectories for the 4-3 covariance method and the 4-5 covariance power method are compared with the response for the

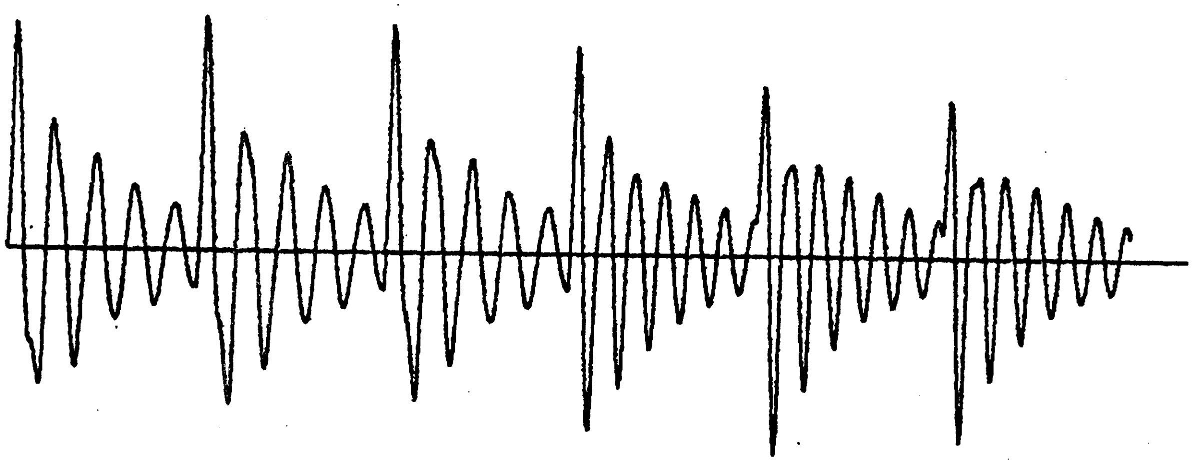

Figure 4.10 Data for 150 Hz Center Frequency Step Change

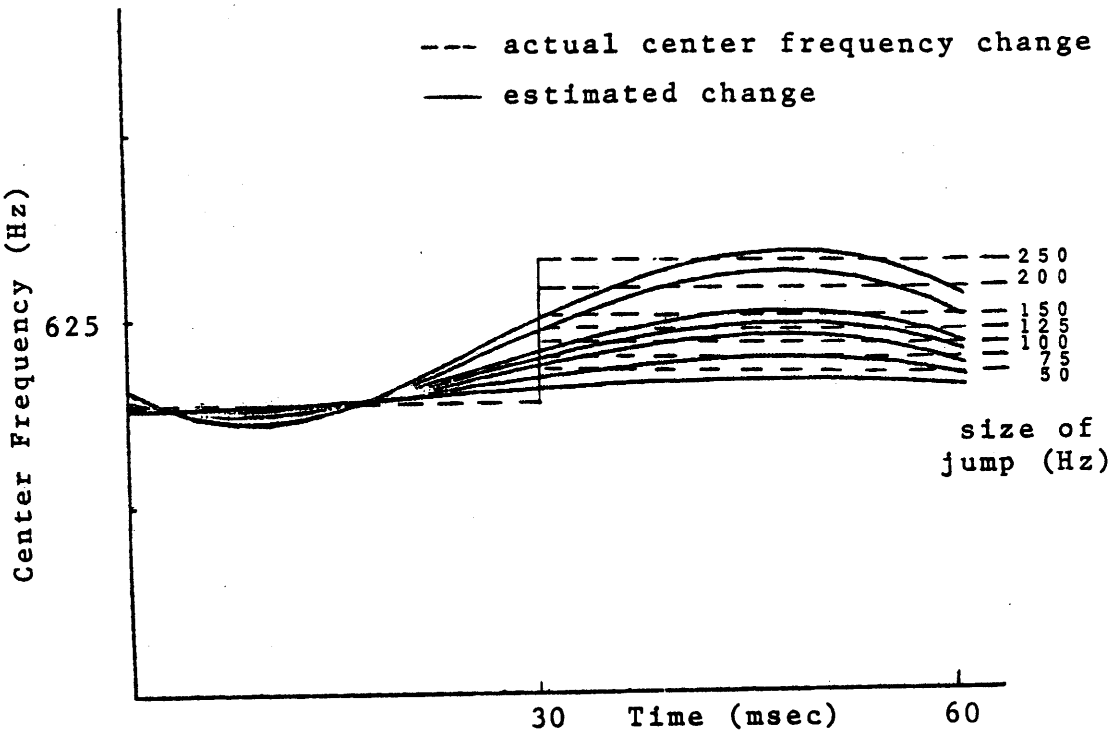

Figure 4.11a Center Frequency Trajectories for 4-3 Covariance Power Filter

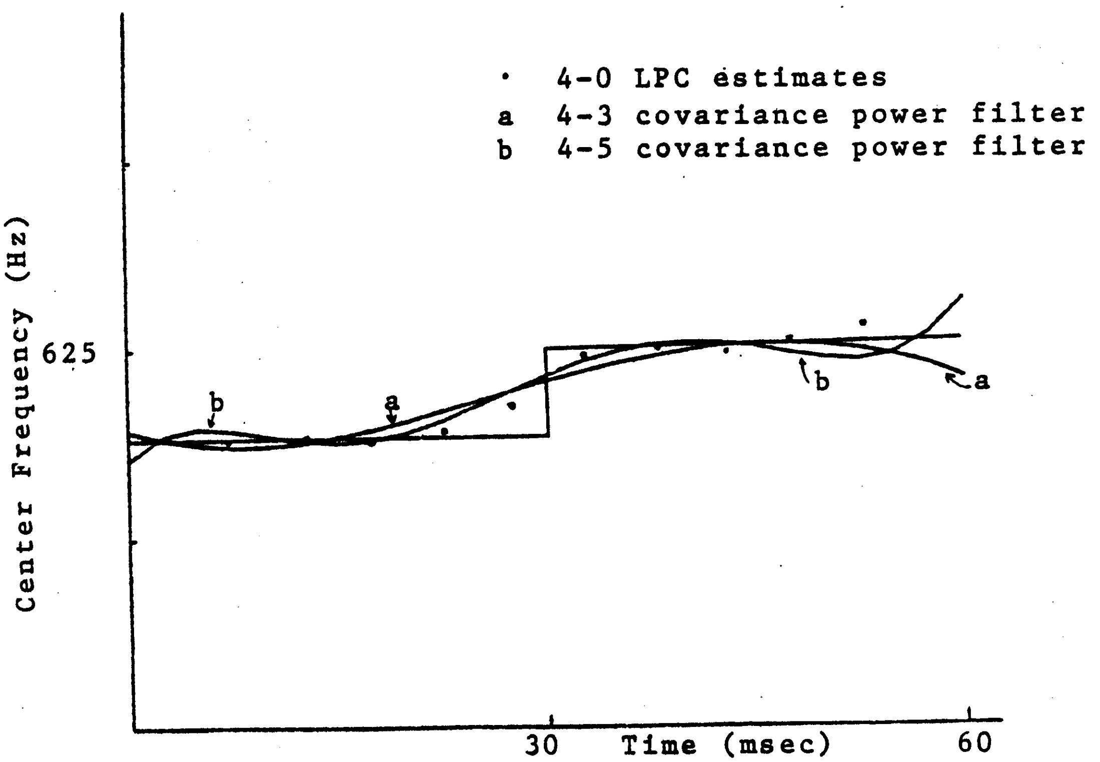

Figure 4.11b Center Frequency Trajectories for 150 Hz Jump

traditional LPC covariance method (the 4-0 covariance method) in figure 4.12b. The LPC method used intervals of 15 msec (150 data points) to estimate the coefficients. The starting location of the analysis interval was shifted by 5 msec for each successive LPC case, so that there was some overlap of the data on each interval. The overlap effectively smoothed the pole trajectory for the standard LPC method. The center frequency of the pole for each interval is plotted at the time corresponding to the center of the interval. Windowing was not used for any of these methods.

From the graph it can be seen that traditional LPC has a response time that is faster than that of the 4-3 covariance power method and is similar to that of the 4-5 covariance power method. However the 4-5 power method shows some irregularities at both ends of the interval.

Since the method is approximately homogeneous its response to the size of the jumps, the next set of cases were developed to see if the method is additive (and hence linear). Specifically, we have tested to see if the response to two different jumps in one interval is the same as the sum of the responses to each jump taken separately in the same interval. The sample case shown in figure 4.12 has the same initial poles as given for the sample case of figure 4.11. However the data interval is 1000 points (100 msec) and the first pole changes from a center frequency of 475 to 575 Hz at data point 450 and then from 575 to 675 Hz at data point 550. The pole angle trajectory for the 4-4 covariance power method is shown in figure 4.12a.

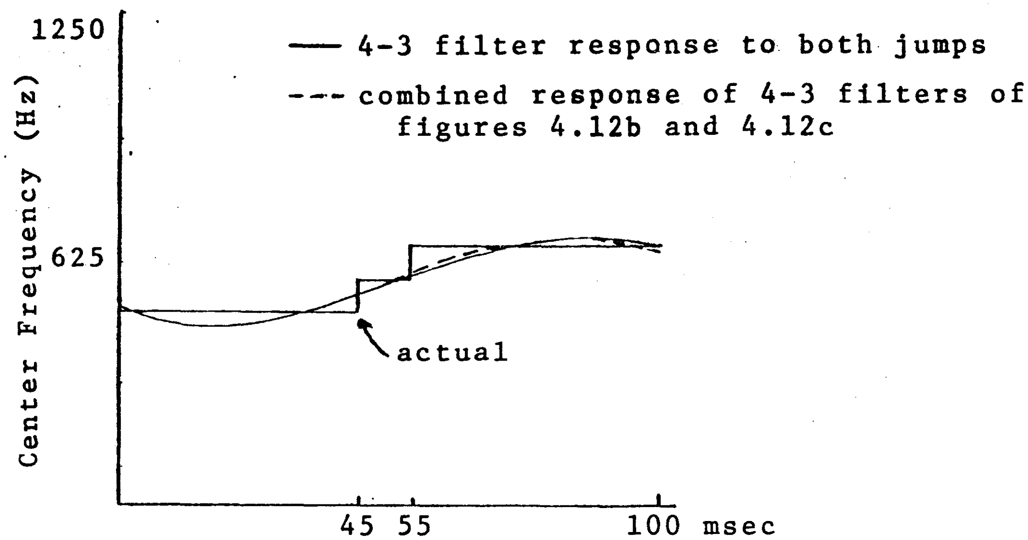

Figure 4.12a Response of 4-3 Covariance Power Filter

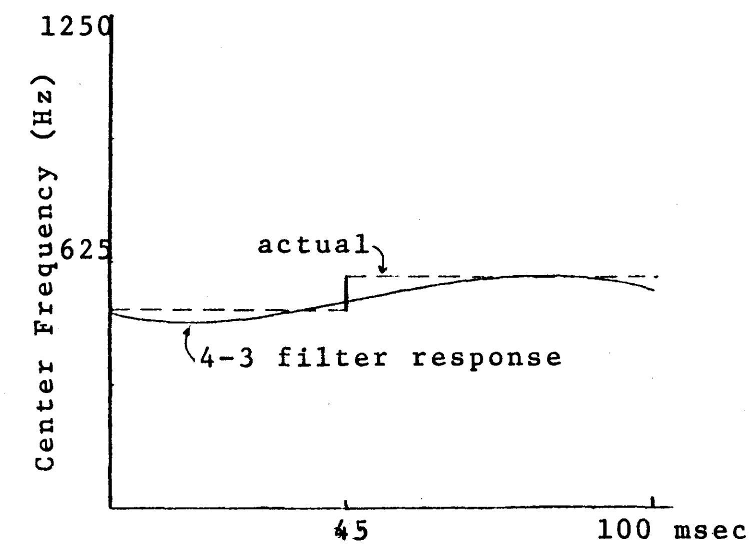

Figure 4.12b Response of 4-3 Filter

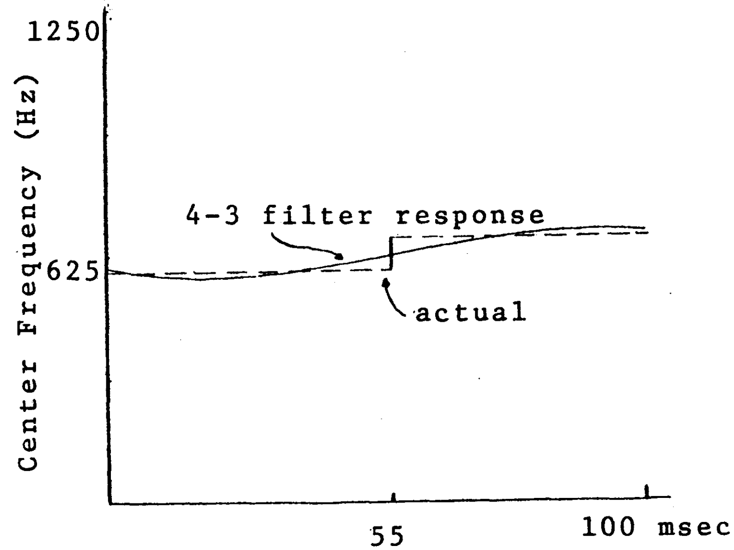

Figure 4.12c Response of 4-3 Filter

The response for the same method for only the pole jump from 475 to 575 Hz at step 450 is shown in figure 4.12b, while the response for the method for the pole center frequency starting at 575 Hz and and then changing at step 550 to 675 Hz is shown in figure 4.12c. Combining these two responses, the total response given by the dotted line of figure 4.12a is obtained. The similarity between the response of the 4-4 power filter for both jumps and the sum of the responses of the filters for each jump is remarkable.

A very similar test case is shown in figure 4.13, where the only difference is that the changes in the center frequency of the pole occur at step 300 and step 700. Again, there is very little difference between the 4-4 power method response for both jumps and the combined response of the filters as shown in figure 4.13a.

These test cases would indicate that the method can be thought of as acting like a linear low pass filter in response to changes in the location of the poles. The method tends to smear abrupt changes in the pole locations, but it should react well to small or slowly-varying changes.

An estimate of the frequency response of the method's "low-pass" action is given by the unit pulse frequency response. Since we have the "step responses" of the system for the 4-3 and 4-5 covariance power filters as shown in figure 4.11, we can find the "unit pulse response" by passing the step response through a $(1-z^{-1})$ filter (i.e., we are taking the first difference of the sequence containing

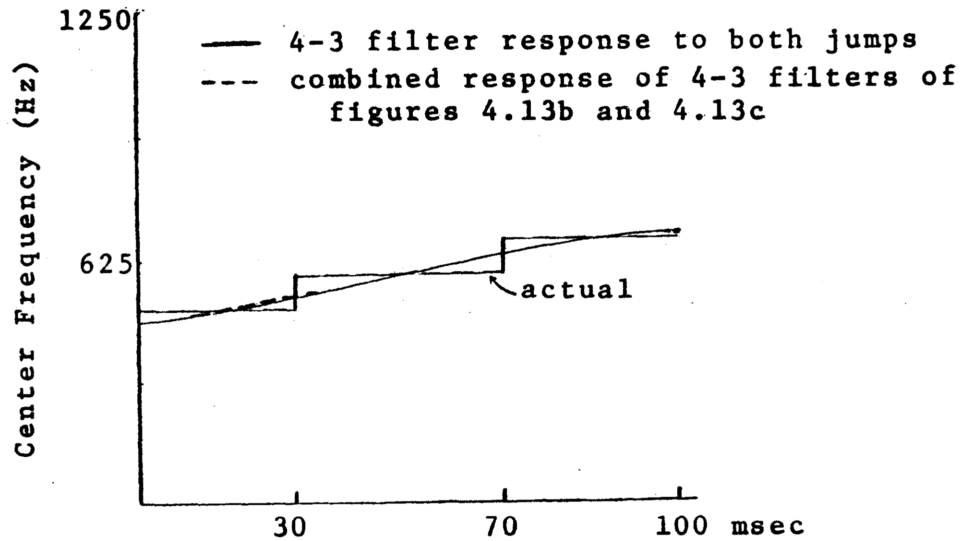

Figure 4.13a Response of 4-3 Covariance Power Filter

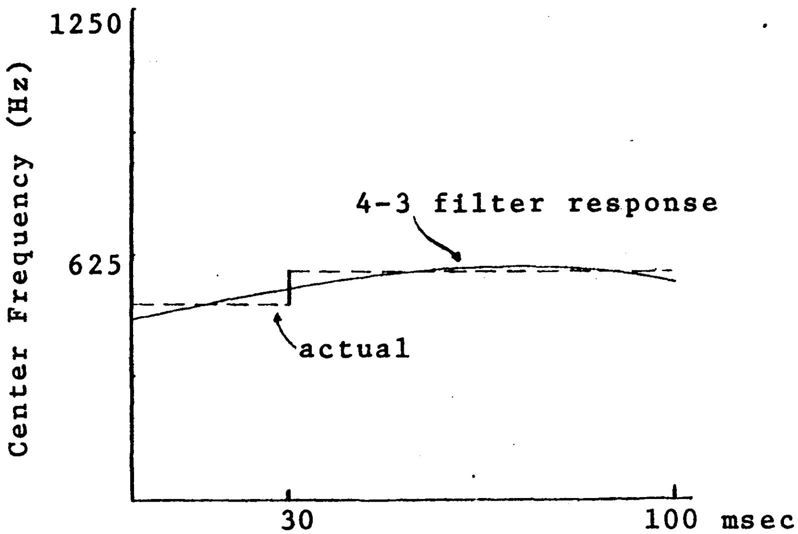

Figure 4.13b Response of 4-3 Filter

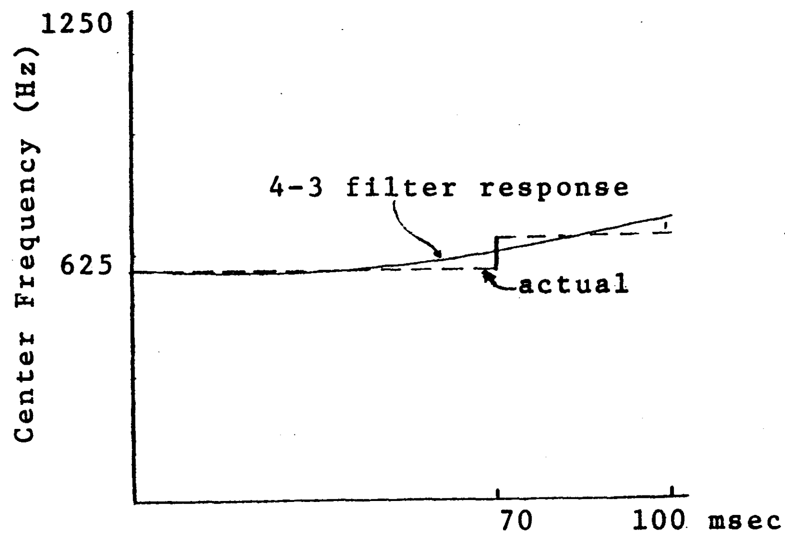

Figure 4.13c Response of 4-3 Filter

the 600 center frequency values). Since the time between center frequency values is .1 msec, the output of the $(1-z^{-1})$ filter can be thought of as the response of the method to a unit pulse of width .1 msec.

By taking the discrete Fourier transform of the unit pulse response, we obtain the unit pulse frequency response. Because the unit pulse is so narrow, it represents a useful approximation to the frequency response of the system. The unit pulse frequency response for the 4-3 covariance power filter is shown in figure 4.14a and the response for a 4-5 covariance power filter is shown in 4.14b. A comparison of the two responses for the frequency range of 0-2000 Hz is given in figure 4.14c (for which the frequency response curves have been smoothed). As we would expect from the "step response" of the two methods, the 4-5 method has a better unit pulse frequency response, that is, it has more high frequency content. Therefore it should be able to track changing center frequencies more accurately than the 4-3 method, because it has a higher "cut-off" frequency.

The unit pulse frequency response for the regular LPC method is not shown because there are not enough sampled values of the center frequency "step response" to obtain the unit pulse response. However, LPC tracked the step change of the center frequency slightly better than the 4-5 covariance power method did (as shown in figure 4.11). Therefore, it should have a unit pulse frequency response that is similar to the 4-5 covariance frequency response,

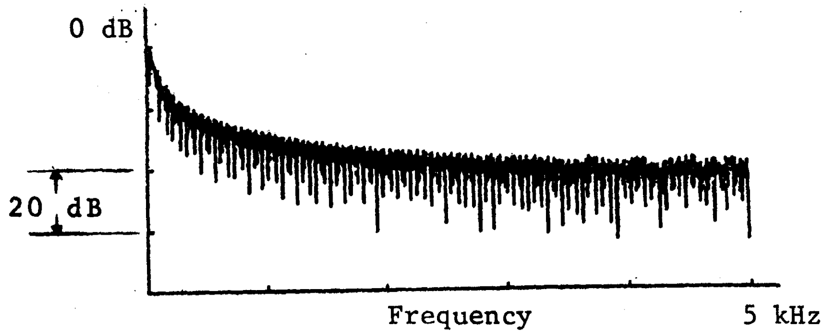

Figure 4.14a Unit Pulse Frequency Response
4-3 Covariance Power Filter

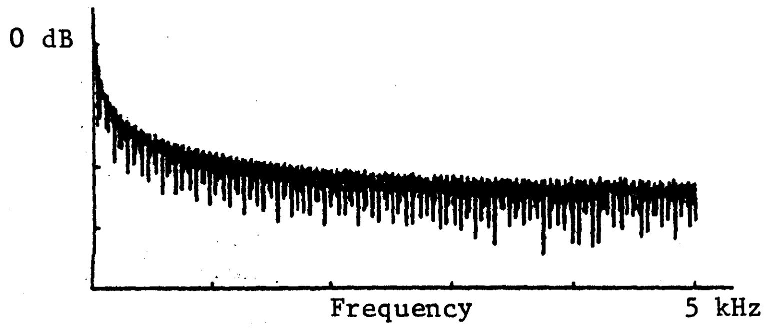

Figure 4.14b Unit Pulse Frequency Response
4-5 Covariance Power Filter

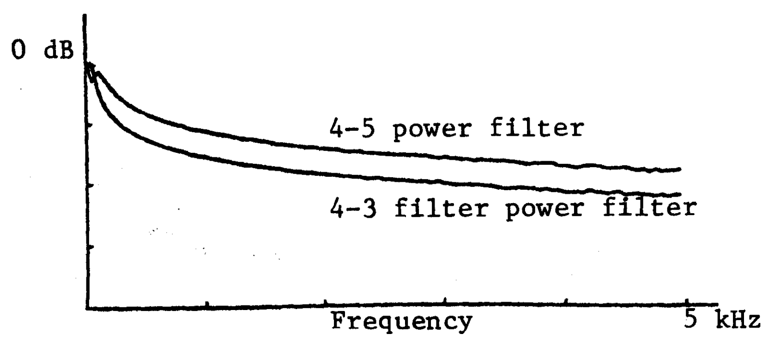

Figure 4.14c Comparison of Smoothed Unit
Pulse Frequency Responses

but the response should have a slightly higher "cut-off frequency".

The next set of test cases were used to evaluate the filter's ability to represent slowly varying changes. The same sample case used in the previous examples was used, however the first pole changed linearly from 475 to 675 Hz over a variable but prescribed time interval. These test cases were 2000 data steps in length and the 4-3 covariance power method was used for analysis. Figure 4.15a shows the pole angle response for a step change of 200 Hz and figure 4.15b shows the response for a linear change of 200 Hz over 200 steps (change begins at step 900 and ends at step 1100). The slope of the change is 10 Hz/msec. The plot of figure 4.15c is for a linear change of 200 Hz beginning at step 700 and ending at step 1300 (slope of 3.33 Hz/sec). The pole trajectory is nearly the same for all three cases (indicating that the changes are still beyond the "cut-off frequency" of the system), however only for the last case does the response follow the change well. The response for the change of 200 Hz over 1000 steps (starting at step 500 and ending at step 1500, which a slope of 2 Hz/msec) is indicated in figure 4.15d. For this example the response matches the slope well. Another test case was created with the same pole trajectory slope, to see if the method could consistently respond well to this slope value. For this example, the pole changed from 475 to 595 Hz (a smaller jump) over 600 time steps (a smaller time interval). The response is shown in figure 4.16a. The method matches the slope of 2 Hz/msec similarly for both of these cases.

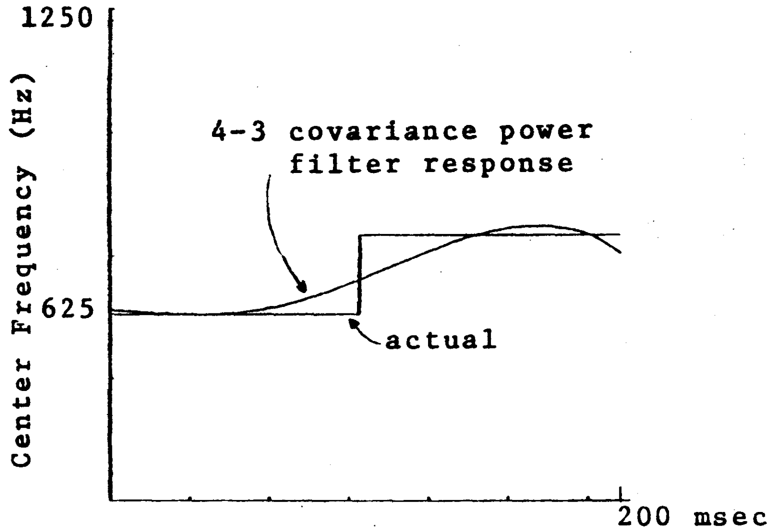

Figure 4.15a Step Change of 200 Hz

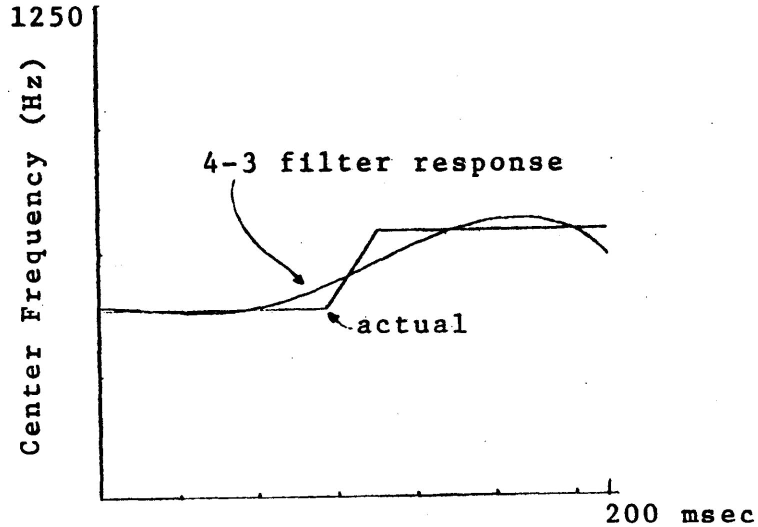

Figure 4.15b Slope of 10 Hz/msec

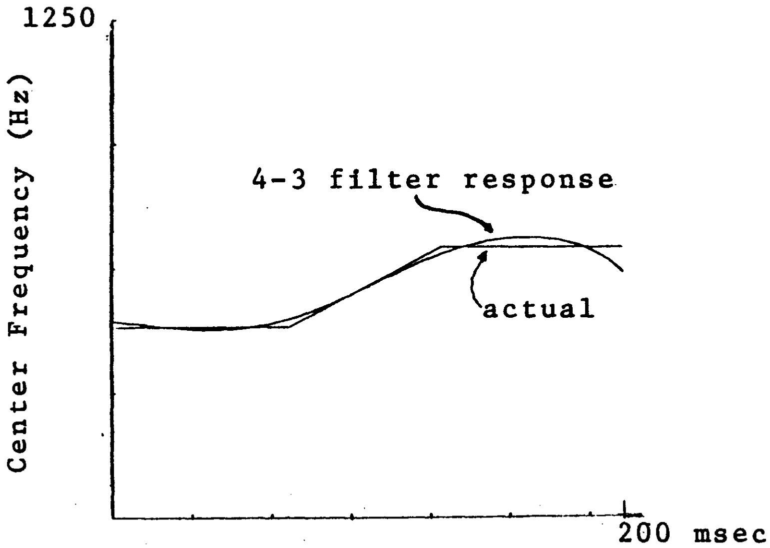

Figure 4.15c Slope of 3.33 Hz/msec

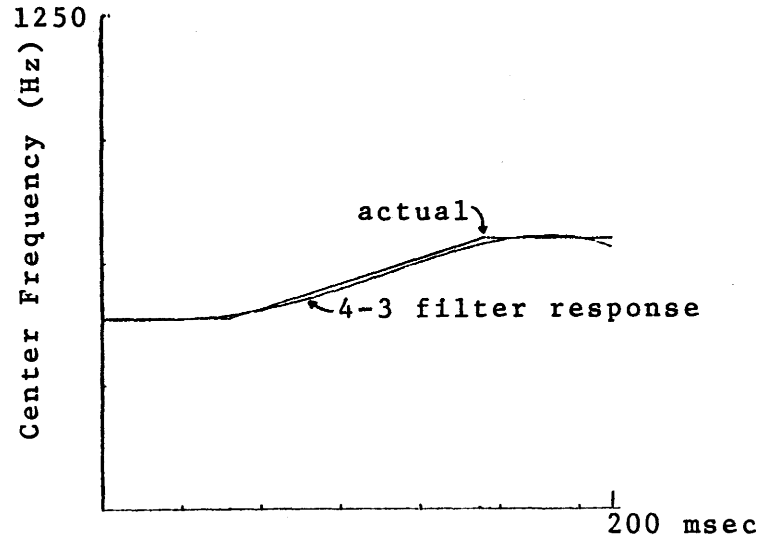

Figure 4.15d Slope of 2 Hz/msec

Figure 4.15 Center Frequency Trajectories

Two additional cases were used to determine if the method could duplicate these slopes using a smaller overall analysis interval of 1200 samples. Figure 4.16b shows the response to a change of 120 Hz starting at step 420 and ending at step 780 (for a slope of 3.33 Hz/msec), and figure 4.16c shows a response to a change of 120 Hz over 600 samples (for a slope of 2 Hz/msec). The response for each case is very similar to the response for the same slope shown earlier. The only significant difference is the time offset of the response slope of the pole angle trajectory for the 2 Hz/msec case.

The conclusion is that time-varying linear prediction can handle linearly changing poles very well if the slope is small. For larger slopes the variation of the pole tends to be smeared over a larger interval. This supports the studies discussed earlier in this section in which we displayed results that indicated the method acted as a low-pass filter. Evidently, the higher slope changes are beyond the cut-off frequency of the method, yielding the same estimated pole trajectory as for an abrupt step change.

From these synthetic test cases, it has been decided that the covariance power method without windowing is probably the best method for analysis. It was shown for at least one example with a 4 pole filter, that the method acts as a low pass filter with respect to step changes in the pole locations of the generating filter. In addition for small linear changes in the poles with respect to time, the time-varying method can duplicate the actual

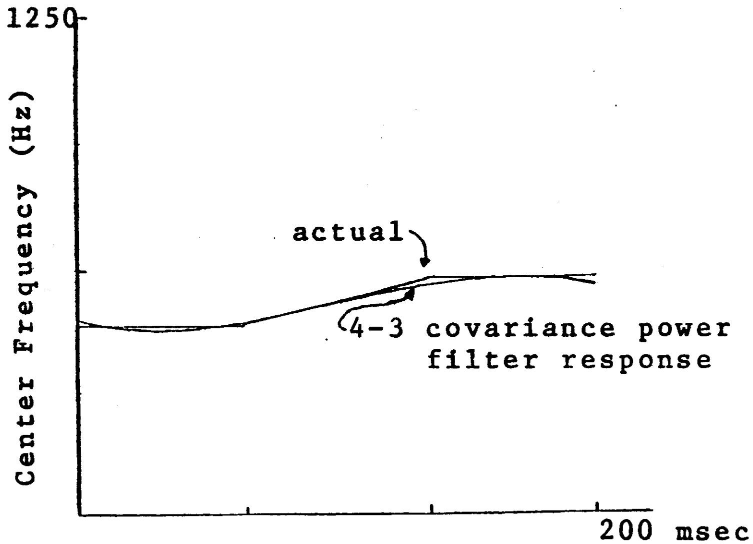

Figure 4.16a Slope of 2 Hz/msec

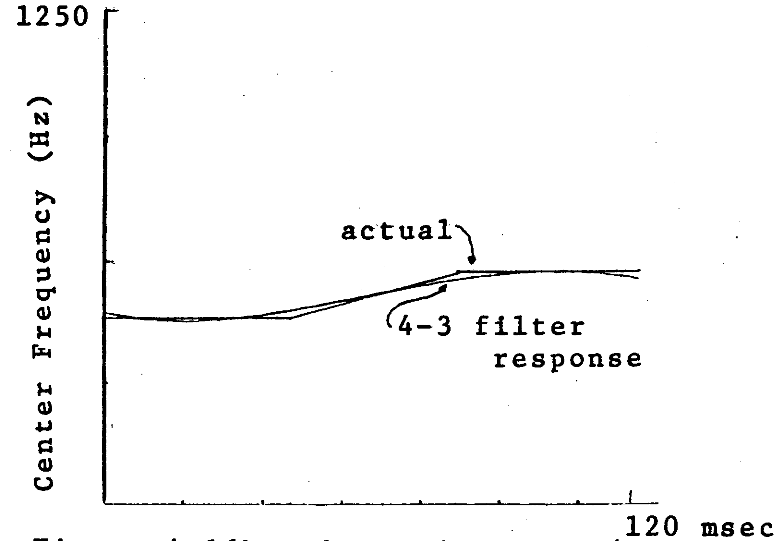

Figure 4.16b Slope of 3.33 Hz/msec

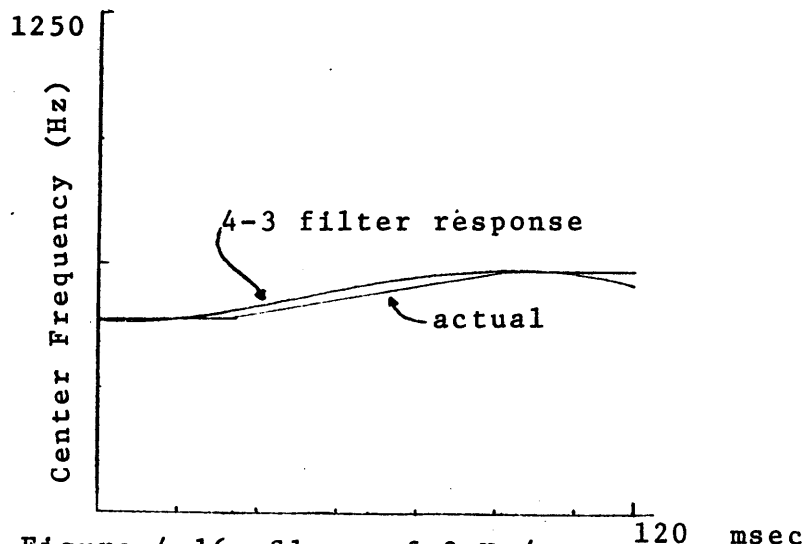

Figure 4.16c Slope of 2 Hz/msec

Figure 4.16 Center Frequency Trajectories pole trajectories very well. It is hoped that the results of this chapter not only shed light on time-varying LPC, but also provide some tools and perspectives for gaining insight into time-varying modelling methods.

CHAPTER V

EXPERIMENTAL RESULTS FOR A SPEECH EXAMPLE

In the last chapter, we presented the results obtained by applying time-varying LPC to synthetic test cases. In this chapter, we will give an example of the application of time-varying LPC to a nonstationary speech waveform. The performance of the method will be examined in depth so that its characteristics can be better determined. In order to evaluate time-varying LPC, its performance will be compared with the results obtained with regular LPC, when it is applied to much smaller "quasi-stationary" segments of the speech waveform.

Several different methods for evaluating the performance of the filters will be used. The pole trajectories of time-varying LPC will be compared with the poles of the time-invariant filters estimated by regular LPC. The log spectrum of each time-invariant LPC filter will also be compared with the log spectrum of the time-varying filter evaluated at the time corresponding to the center of each of the analysis intervals used for regular LPC. As a measure of how well these spectra compare, a log spectral measure given by Markel and Gray [17] and Turner and Dickinson [18] will be used. In addition, the impulse responses of both regular and time-varying LPC will be compared with the original speech data.

The nonstationary speech waveform that was used for this example is shown in figure 5.1a. It contains 1600 data points, which corresponds to a time of .16 sec, since it was sampled at a rate of 10,000 Hz. In order to estimate the spectral properties of the vocal tract, the waveform was pre-emphasized by a simple one-zero filter of the form $1-\mu z^{-1}$ to remove the glottal effects, as suggested by Markel and Gray [3]. For this example, the value of μ was .95 (the value is not critical to the results, i.e. any value between .9 and 1 could be used). The pre-emphasized waveform is shown in figure 5.1b.

Markel and Gray [3] state that a reasonable value for the order of a prediction filter for speech data is usually between 12 and 16. For this example, we have chosen a value of p=12. The time-varying model that was used was a 12-5 power series filter. (Because of the results of the last chapter, the use of a Fourier series seemed repetitious and unnecessary.) The time-varying LPC analysis was performed on an interval containing the first 1500 samples.

For regular LPC, a 12 pole filter was used and the length of each analysis interval was 200 samples. The center of the interval was shifted by 150 samples for each successive LPC analysis, resulting in some overlap of the data contained in each interval. The only exception to this was for the first analysis interval, which contained only the first 100 data points. The second interval was 200 points in length and had its center at time step 150. The very last interval was also 200 points in length and started at time step 1400. Therefore it contained samples of the speech

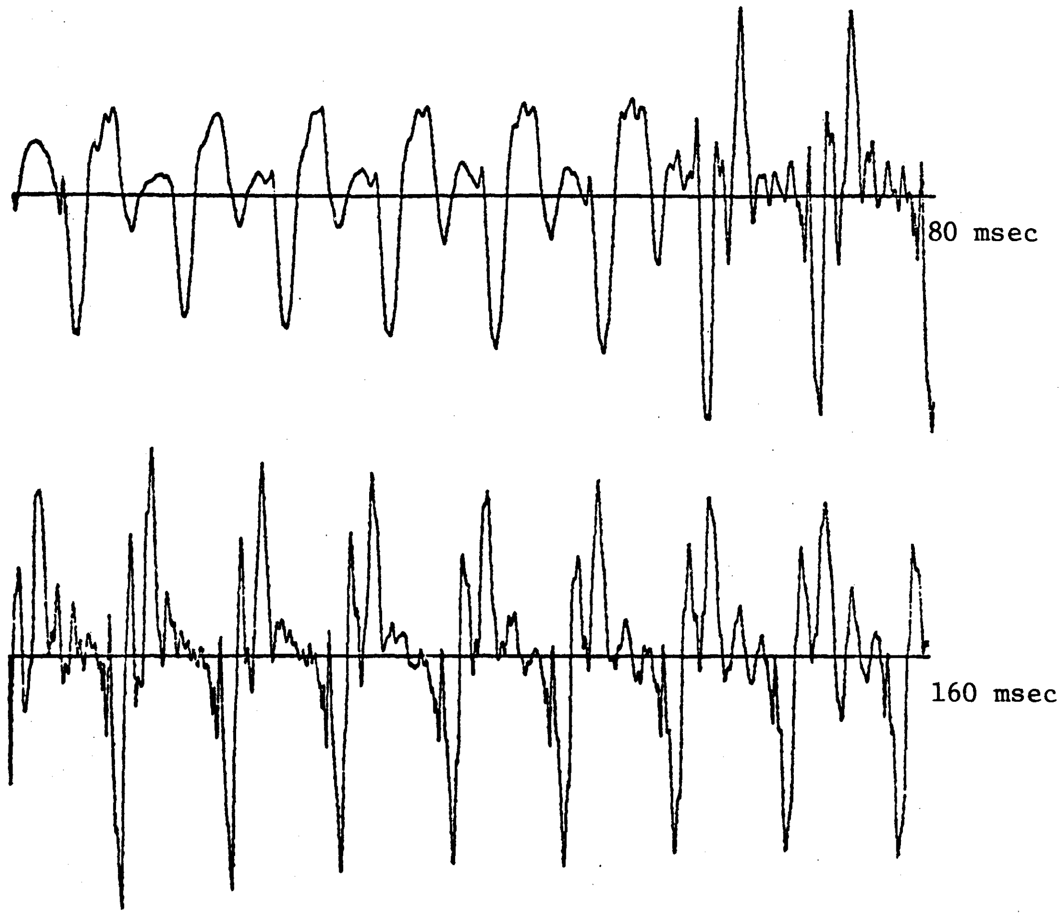

Figure 5.1a Nonstationary Speech Waveform

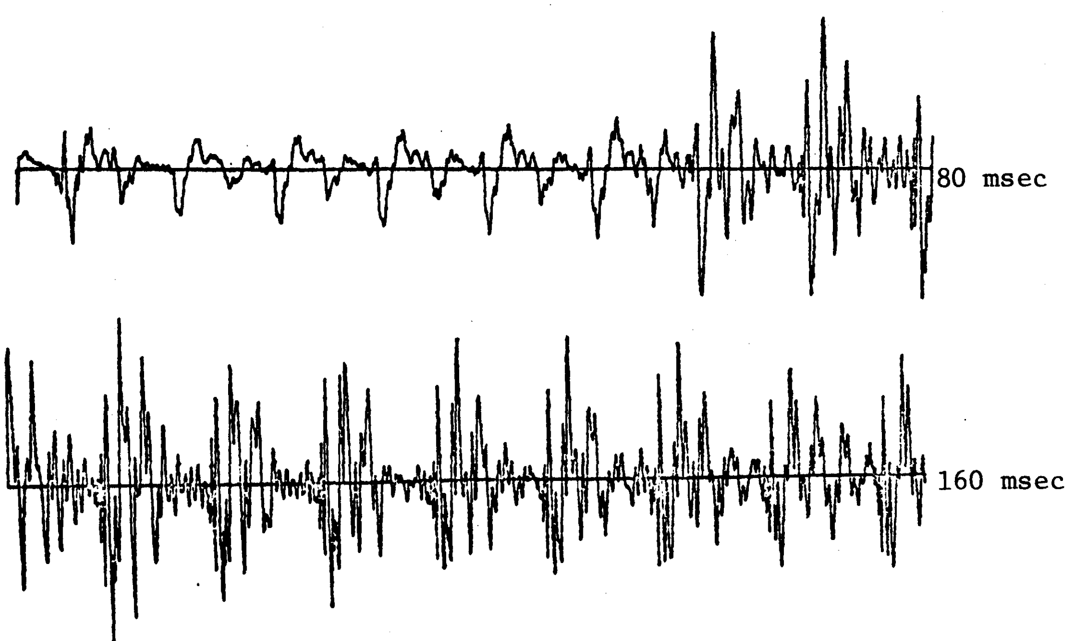

Figure 5.1b Pre-emphasized Version of Speech Waveform of Figure 5.1c

waveform for the time 1500 to 1599 that were not used for the time-varying LPC method.

For the regular LPC analysis, the covariance method was used, both with and without windowing the data. The results for both methods were so similar that only the covariance LPC method without windowing will be compared with the time-varying LPC method.

The pole trajectories for the covariance power series method both with and without windowing the data are shown in figure 5.2. This illustrates dramatically the effect of windowing, because there are poles of the filter for the windowed data that are outside the unit circle. For a time-invariant filter, this would mean that the filter was unstable. For a time-varying filter, this is not necessarily true. However the few time-varying filters we have examined that have had some poles outside the unit circle have had impulse responses that usually remain bounded but excessively large. In general, the time-varying filter with poles outside the unit circle would seem to be of no practical value.

The pole trajectories for the 12-5 autocorrelation power series filter are shown in figure 5.3. Again, the autocorrelation filter for the windowed data has poles outside the unit circle. The results of the autocorrelation method (without windowing) agrees favorably with that of the covariance method. The most significant differences occur at each end of the interval (as we would expect from our discussion in chapter 4).

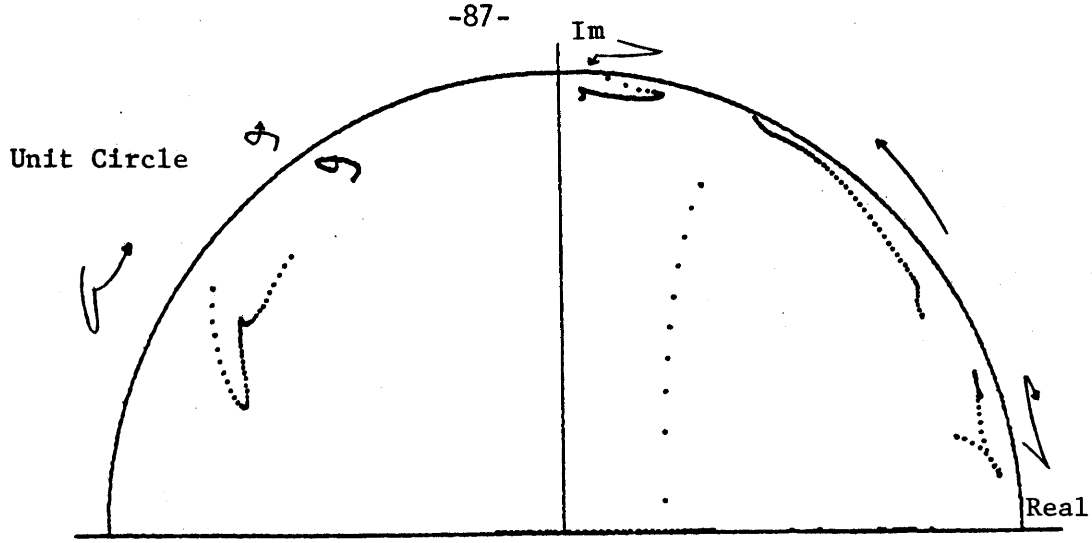

Figure 5.2a Pole Trajectories for 12-5 Covariance Power Filter (data not windowed)

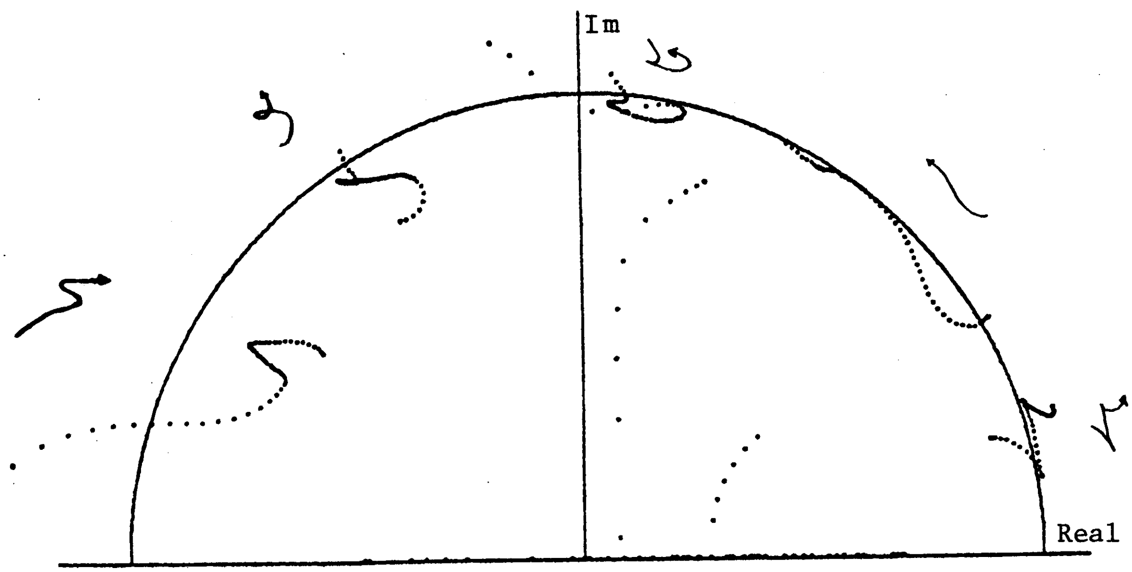

Figure 5.2b Pole Trajectories for 12-5 Covariance Power Filter (data windowed)

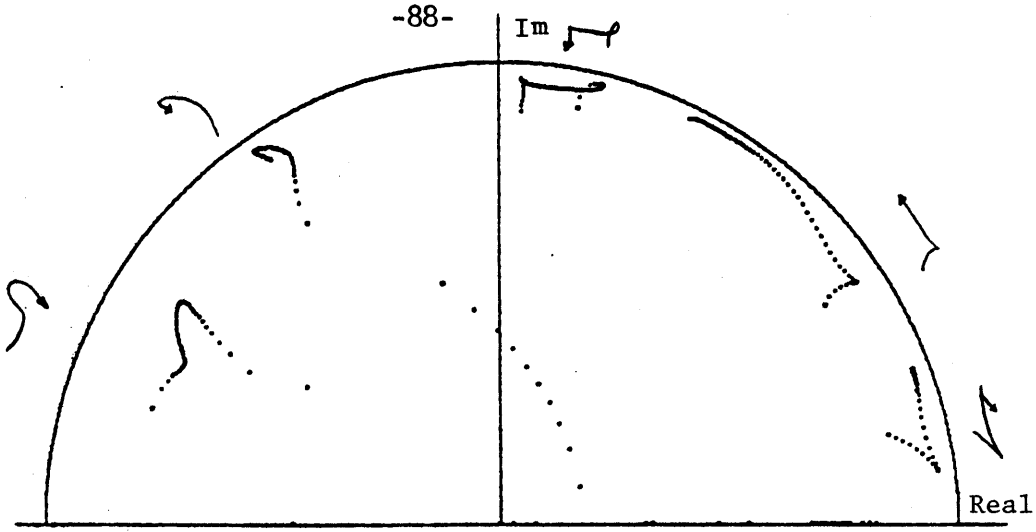

Figure 5.3a Pole Trajectories for 12-5 Correlation Power Filter (data not windowed)

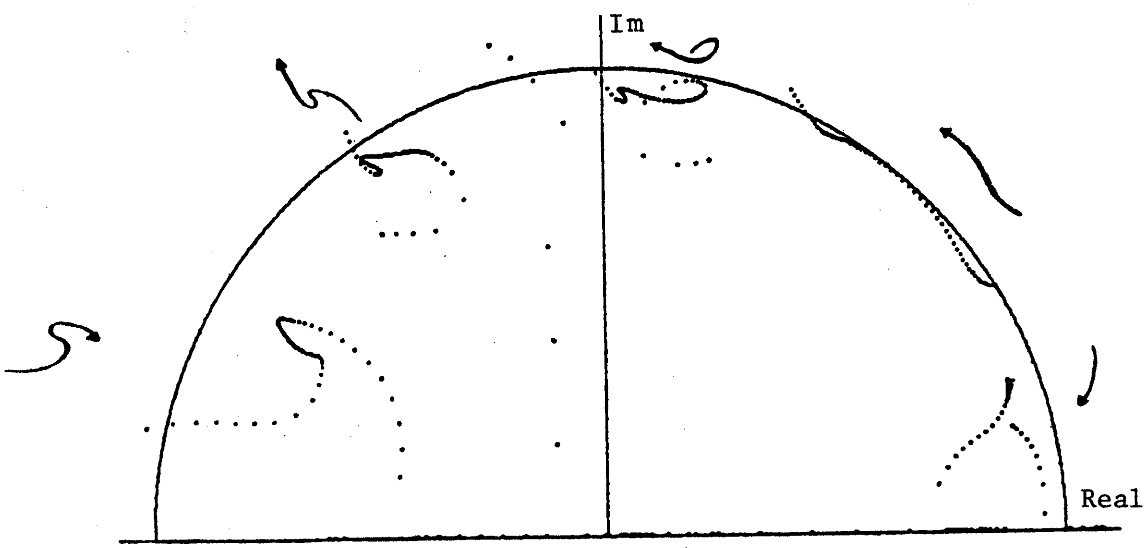

Figure 5.3b Pole Trajectories for 12-5 Correlation Power Filter (data windowed)

This example shows that for any of the time-varying methods developed in this thesis, there is no guarantee that the poles of the filter will remain inside the unit circle. This is a limitation of the time-varying method, but whether it is a serious problem in general practice is not known. Because windowing the data seems to increase the probability that the resulting filter will have poles outside the unit circle, it seems that the data should not be windowed. Since the covariance method seems better justified analytically than the autocorrelation method, the covariance power method (without windowing) will be used for comparison with regular LPC.

For the covariance power method, it can be seen that there are only 5 sets of complex poles over much of the interval. The other two poles were generally real. This was also true occasionally for the time-invariant filters determined using regular LPC. For comparison purposes, only the five sets of poles that were always complex will be compared with the time-invariant LPC poles.

The center frequency trajectories of the complex poles are shown in figure 5.4. The radius trajectories for each pole are shown in figure 5.5. The center frequencies and radii of the poles for the time-invariant filters are also shown on the figures. The values are plotted at the time corresponding to the center of the analysis interval.

The trajectories of the center frequencies for both methods

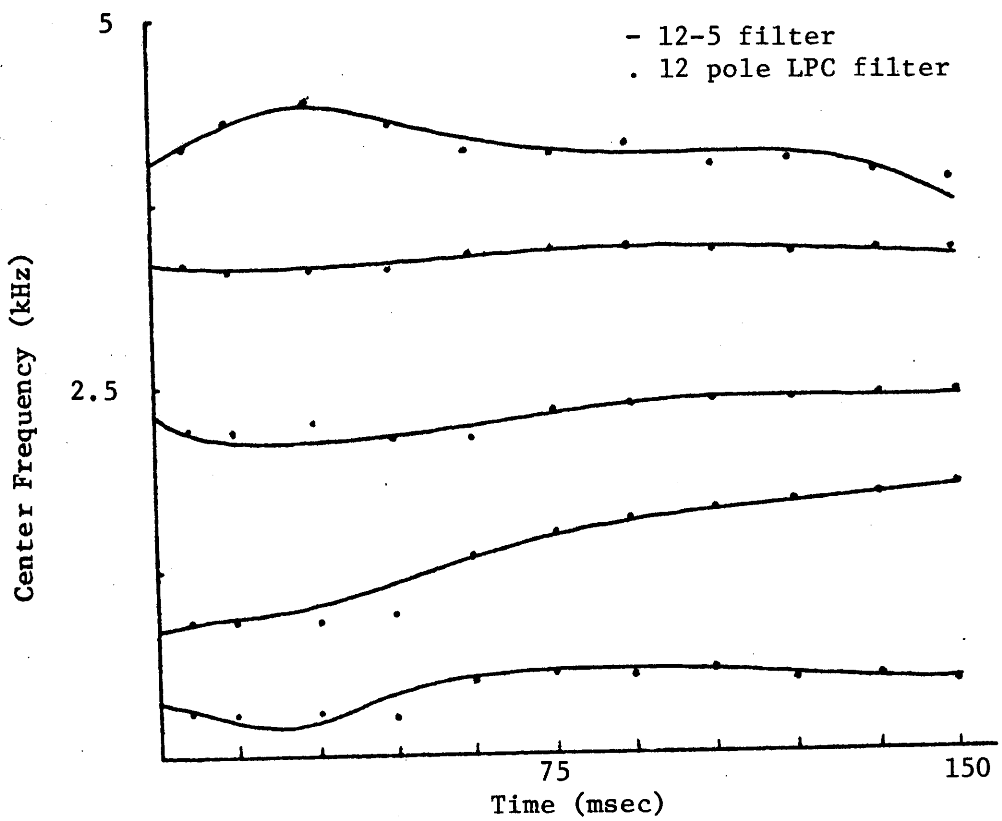

Figure 5.4 Center Frequency Trajectories for 12-5 Covariance Power Filter and 12 Pole LPC Filters.

- 12-5 Covariance power filter
. LPC

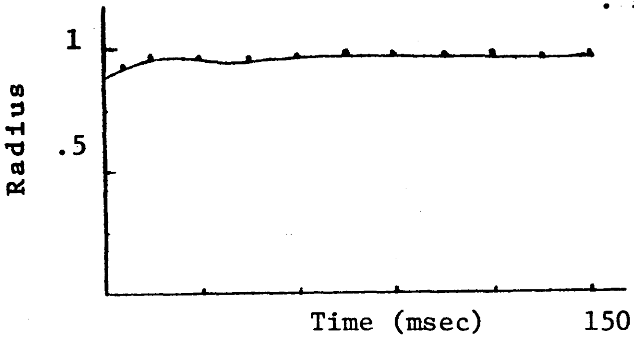

Figure 5.5a First Pole

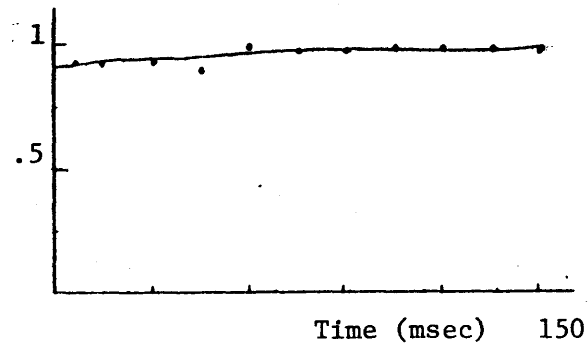

Figure 5.5b Second Pole

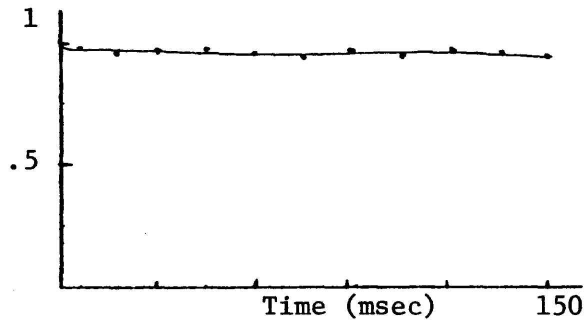

Figure 5.5c Third Pole

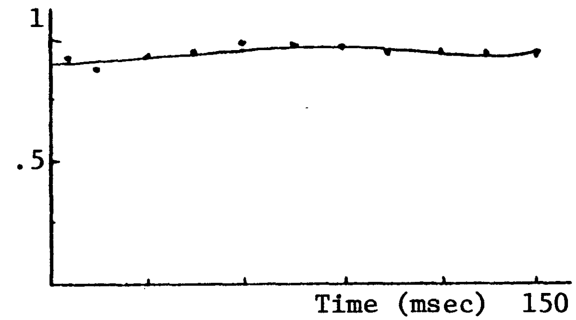

Figure 5.5d Fourth Pole

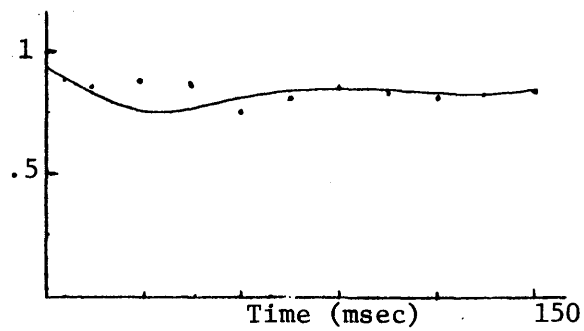

Figure 5.5e Fifth Pole

Figure 5.5 Radius Trajectories

agree favorably. The main deviations between the time-varying method and regular LPC occur in the first and second poles in the time interval of 150 to 600, where the "low-pass" nature of the time-varying LPC method is most evident. But after time step 600 the correspondence between the two methods is very good. The time-varying method can be seen to be "smoothed" values of the center frequency locations of regular LPC. The radius trajectories of the poles agree fairly well, except for the fifth pole. It is interesting to note that the center frequency trajectory of the fifth pole matches very well, while the radius trajectory does not. The radius trajectory deviations seem to be a result of the "low-pass" nature of the time-varying LPC method.

Next we will compare the log spectra of the all-pole time-invariant and time-varying filters with log spectra of the speech signal. The spectra will be compared because LPC can be thought of as attempting to match the spectral envelope speech with the spectrum of the all-pole filter. This is discussed in detail in [3,6]. The spectrum $X(e^{j\omega})$ is found by taking the discrete Fourier transform (DFT) of the sequence $(x(n), n=0,1,\ldots,N-1)$ [3]. To obtain better frequency resolution, zeros can be appended to the end of the sequence. The log spectrum LM(X) is given by

$$LM(X) = 10 \log_{10} |X(e^{j\omega})|^2 \qquad (5.1)$$

The speech spectra have been calculated by taking the DFT of the speech samples in the intervals used for the regular LPC analysis.

The spectrum of the all-pole time-invariant filter, $H(z) = 1/A(z)$ (where the filter has p coefficients $(a_i, i=1,2,\ldots,p)$) is found by taking the DFT of the sequence $[1,a_1,a_2,\ldots,a_p]$. The log spectrum LM(H) is given by

$$LM(H) = LM(1/A) = -10 \log_{10}|A(e^{j\omega})|^2 \tag{5.2}$$

For a filter with time-varying coefficients one can only talk about spectrum in an intuitive way. However, when the coefficients vary slowly, the following approach appears to have merit in allowing us to understand the performance of time-varying LPC. Let $(a_i(n), i=0,\ldots,p;\ n=0,\ldots,N-1)$ be the coefficients of the time-varying filter. Then a spectrum can be calculated at time n=k by taking the DFT of the sequence $[1,a_1(k),a_2(k),\ldots,a_p(k)]$, with the log spectrum $LM(H_k)$ being

$$LM(H_k) = LM(1/A_k) = -10 \log_{10}|A_k(e^{j\omega})|^2 \tag{5.3}$$

(where the subscript k denotes that the coefficients of the time-varying filter have been evaluated at time n=k). The spectra of the time-varying filter have been calculated for the values of n corresponding to the center of each interval used for the regular LPC analysis. Since the pitch period of the excitation function is usually rather large, the concept of a spectrum for the slowly changing time-varying filter is reasonable and provides some insight into the filter's characteristics.

The spectra for the regular LPC and time-varying LPC filters are shown superimposed upon the speech spectra in figures 5.6 and 5.7. For these spectra, the length of the DFT was 1024 points. The spectra have been adjusted so that the largest value is 0 dB.

We shall use a log spectral measure to determine quantitatively the difference between the spectra for both LPC methods [17,18]. Following the derivation given by Turner and Dickinson [18], the RMS log spectral measure, d_2, for the comparison of two all-pole filters ($G/A(z)$ and $G/A'(z)$) is given by

$$(d_2)^2 = \int_{-\pi}^{\pi} \left| \ln(G^2/|A(e^{j\theta})|^2) - \ln(G^2/|A'(e^{j\theta})|^2) \right|^2 \frac{d\theta}{\pi} \qquad (5.4)$$

The Taylor series expansion for ln A(z) (assuming A(z) is stable) is

$$\ln A(z) = -\sum_{k=1}^{\infty} c_k \, z^{-k} \qquad (5.4)$$

with the cepstral coefficients given by

$$c_0 = \ln(G^2) \qquad (5.4)$$

$$c_k = -a_k - \sum_{n=1}^{k-1} c_{k-n} \, a_n \, \frac{(k-n)}{n} \qquad k > 0$$

By applying Parseval's relationship to 5.4, the log spectral measure is

$$(d_2)^2 = \sum_{k=-\infty}^{\infty} (c_k - c_k')^2 \tag{5.5}$$

with $c_k = c_{-k}$. By using only the first p terms and scaling for a dB variation in the power spectrum, the spectral measure SPDIFF is given by

$$\text{SPDIFF} = \left(\frac{10}{\ln 10}\right) \cdot \left[2 \sum_{i=1}^{p} (c_k - c_k')^2\right]^{1/2} \tag{5.6}$$

Markel and Gray [17] have reported that there is a high correlation between SPDIFF and d_2. Turner and Dickinson [18] state that perceptual studies have shown that SPDIFF changes of 2 dB are barely noticeable, but that changes of 3.5 dB are consistently perceptible.

Turner and Dickinson also develop an average SPDIFF for filters with time-varying coefficients. For the examples in this chapter, we want to compare a filter that has constant coefficients $(a_i, i=1,\ldots,p)$ with a filter that has time-varying coefficients $(a_i'(n), i=1,\ldots,p)$, where n is evaluated over an interval of interest (which, for now, we will assume to be [1,L]). For this, the time average spectral difference is

$$\text{AVG SPDIFF} = \left(\frac{10}{\ln 10}\right) \cdot \left[\frac{1}{L} \sum_{n=1}^{L} 2 \sum_{k=1}^{p} (c_k - c_k'(n))^2\right]^{1/2} \tag{5.7}$$

where the cepstral coefficients $c_k(n)$ are calculated from 5.4 using the coefficients $(a_i(n), i=1,\ldots,p)$. This is a measure of the average spectral difference between the time-invariant filter and the time-varying filter over the interval [1,L].

The spectral difference, SPDIFF, between the regular LPC estimated filter and the time-varying LPC filter evaluated at the time corresponding to the center of the regular LPC analysis interval is given in figures 5.6 - 5.8. The time average of the spectral difference between the regular LPC filter and the time-varying filter for all the time steps n in the corresponding regular LPC analysis interval is also listed (i.e., we compute 5.7 for the data interval used in the corresponding LPC analysis). As an indication of how quickly the speech spectrum is changing, the spectral difference between the regular LPC filters for successive analysis intervals is given.

There are large spectral differences between the successive regular LPC time-invariant filters for the comparison times of 50 and 150, 450 and 600, 600 and 750. These are the times where the signal characteristics are changing significantly. The largest average spectral differences between the time-varying LPC filter and the regular LPC time-invariant filters occur at the times of 300 and 450, (as to be expected from the comparison of the pole trajectories in figures 5.4 and 5.5). The values of the average spectral differences were 2.5 and 3.4 respectively, which would

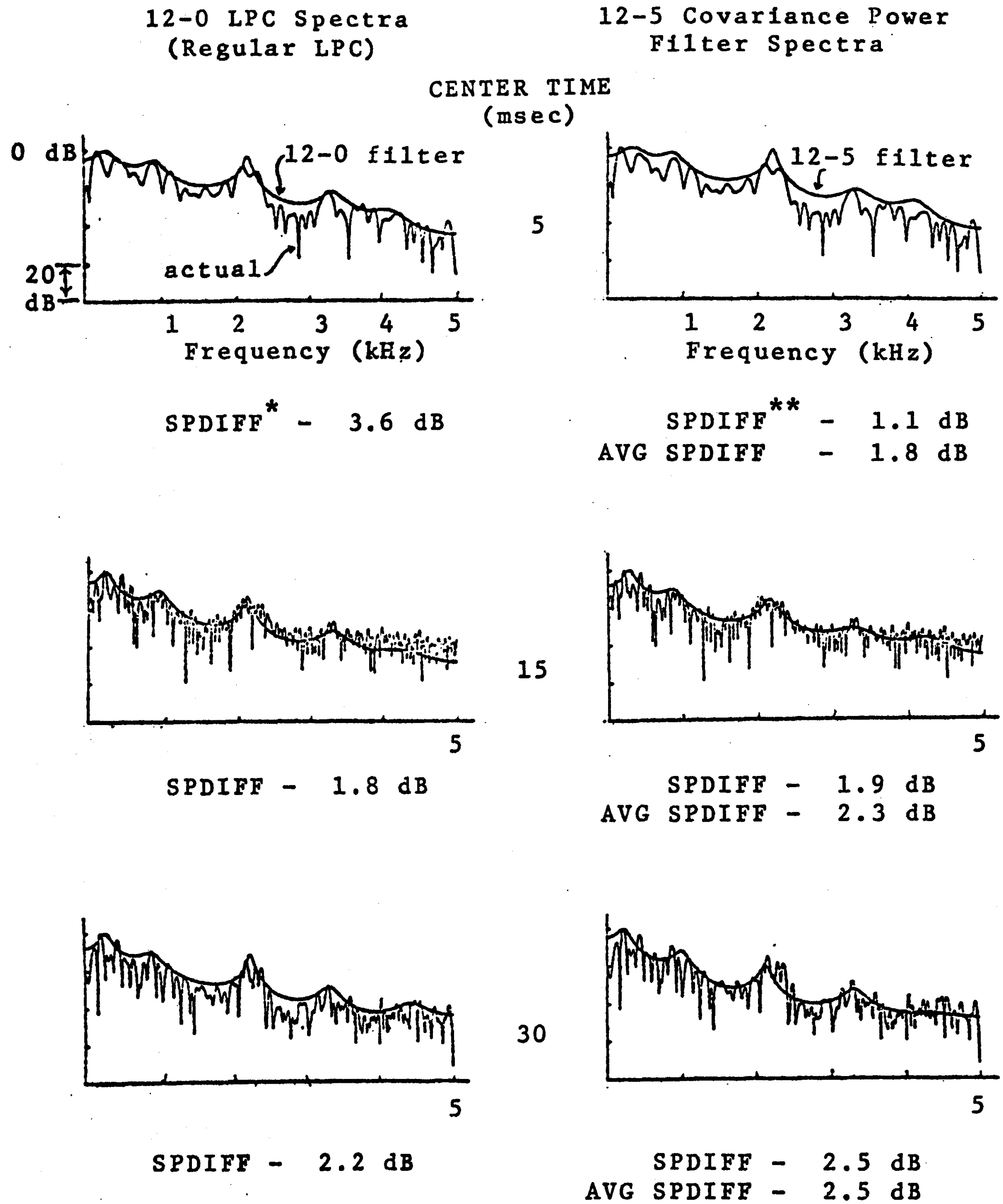

Figure 5.6 Comparison of Actual and Filter Spectra for Pre-emphasized Speech Example

*Difference between 12-0 (regular) LPC spectra for successive center times (i.e., between 5 and 15 msec)

**Difference (and average difference) between 12-0 and 12-5 filter spectra for same center time

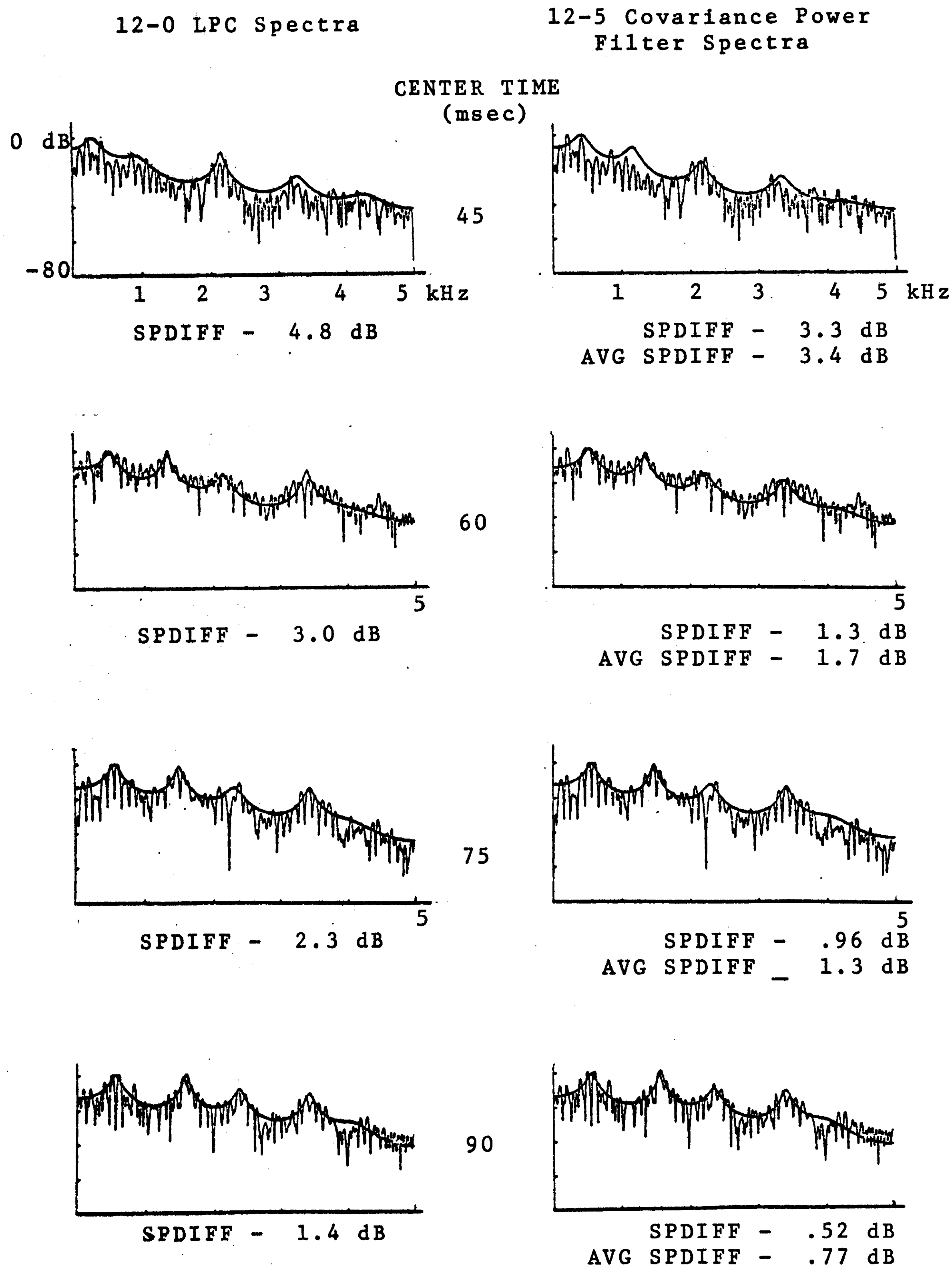

Figure 5.7 Comparison of Actual and Filter Spectra (cont.)

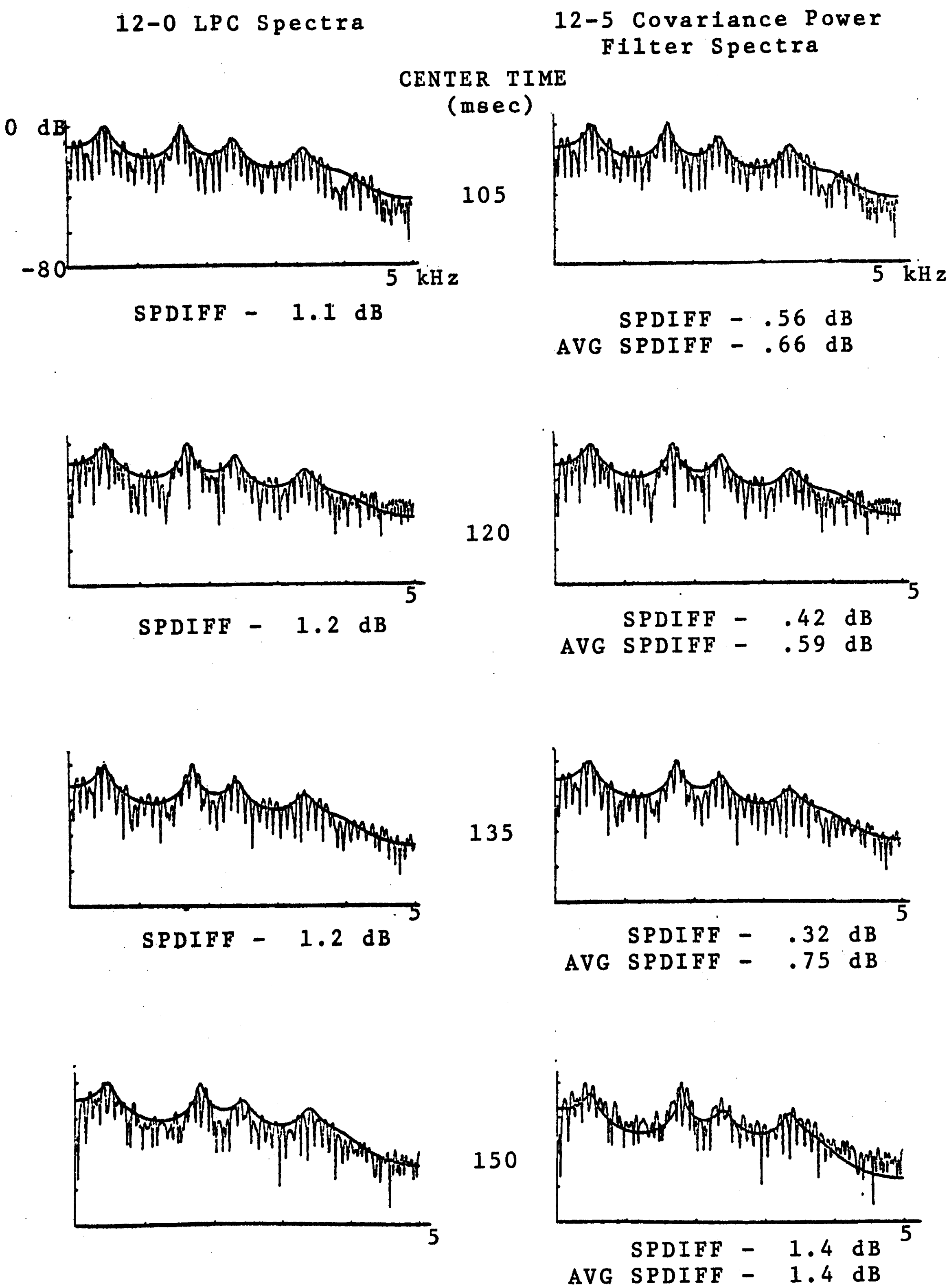

Figure 5.8 Comparison of Actual and Filter Spectra (cont.)

indicate that the differences between the two methods would be perceptible. After time step 600, the average difference between the time-varying spectra and the time-invariant spectra were generally less than the difference between the time-invariant spectra for successive intervals, which would signify that the time-varying method is "tracking" the changing spectra very well.

The relatively large deviation of the time-varying spectrum from the actual speech spectrum for the times around 450 can be explained in part because of the "low-pass" action of the time-varying filter. However the severity of the deviation is probably also due to the unequal energy distribution of the speech signal and of the impulse driving the system. There is much more energy in the latter part of the signal (after time step 600). It was determined by examining the error sequence, e(n), that there was also more energy in the impulses driving the system after time step 600. Therefore the least squares error techniques will produce filters that fit the latter data better. This is especially evident from the center frequency trajectories (fig. 5.4), where it can be seen that the center frequencies of the poles of the time-varying and regular LPC filters compare very well for the time after point 600. The conclusion is that the time-varying filters should match the high energy areas of the nonstationary signal the best. In order to have a relatively good match over all the data in the interval, the energy of the signal or the driving impulses throughout the entire interval should be approximately equal.

To test this hypothesis, the original pre-emphasized signal was modified and used for analysis. For the first modification, the initial portion of the signal [0,574] was multiplied by 2 (the data for [575-1499] was not changed) so that the average energy of the signal was approximately equal throughout the entire interval. The modified signal was analyzed by time-varying LPC and it was found that the resulting center frequency trajectories matched the LPC center frequency estimates somewhat better than for the time-varying LPC analysis of the original pre-emphasized signal.

By examining the error sequence, e(n), it was evident that the driving impulses still had more energy for the latter portion of the signal. Therefore to equalize the input energy over the interval, the initial part [0,574] of the original signal was multiplied by 4, which resulted in the modified signal of figure 5.9a.

Using this signal for analysis by the 12-5 covariance power method resulted in the pole trajectories of 5.9b. The center frequency trajectories are shown in figure 5.10. These trajectories matched the LPC estimated center frequency values for the interval of [0,600] much better than the original 12-5 covariance power filter did.

The spectra for the 12-5 covariance power filter for different time steps are shown in figure 5.11. The values of the average spectral difference, AVG SPDIFF, between the spectra of the regular LPC filter (the filter that was estimated using the original data) and the time-varying 12-5 power filter for each LPC analysis interval

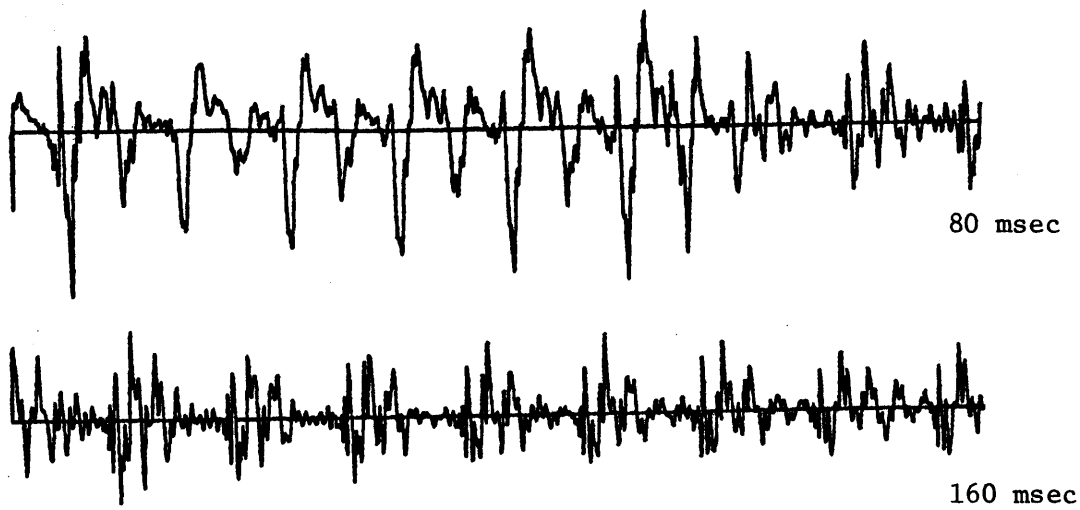

Figure 5.9a Modified Pre-emphasized Speech Signal
(first 57.5 msec multiplied by 4)

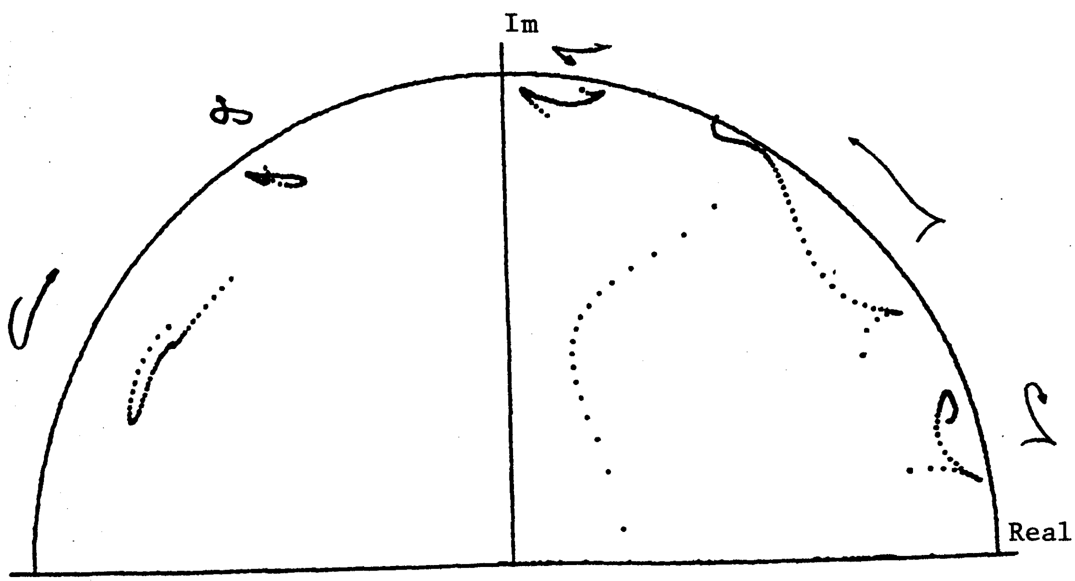

Figure 5.9b Pole Trajectories for 12-5 Covariance Power Filter for Modified Data Shown Above.
(data not windowed)

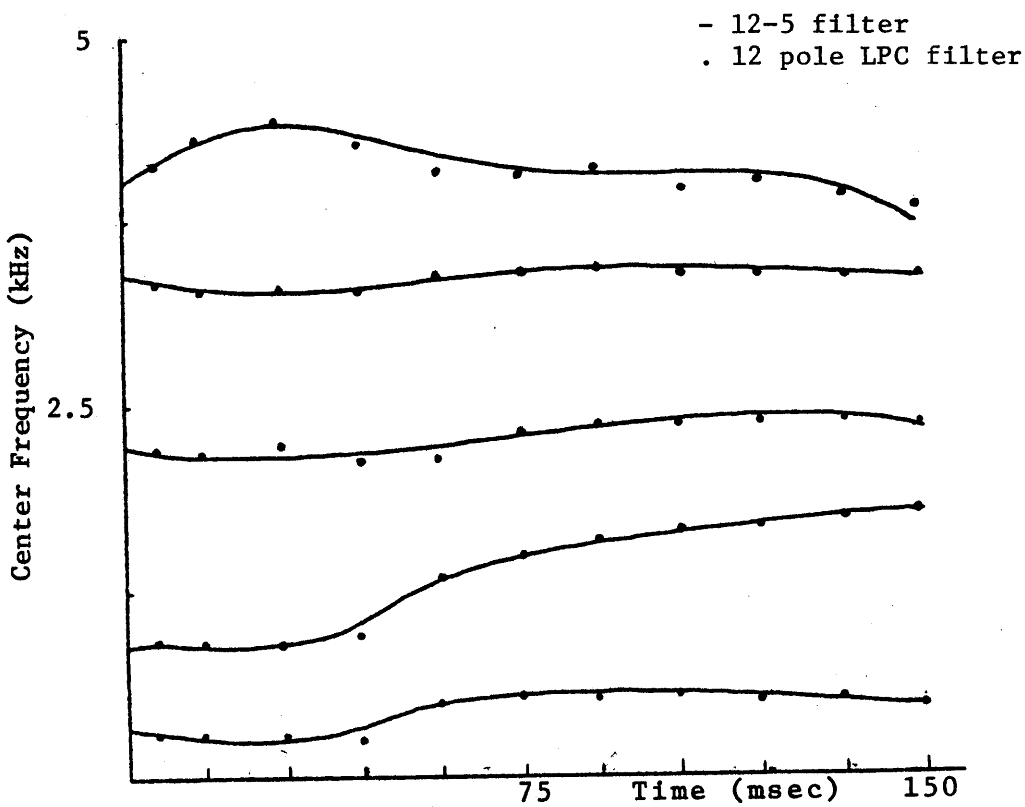

Figure 5.10 Center Frequency Trajectories for 12-5 Covariance Power Filter (for modified data) and 12 Pole LPC Filters (for original data).

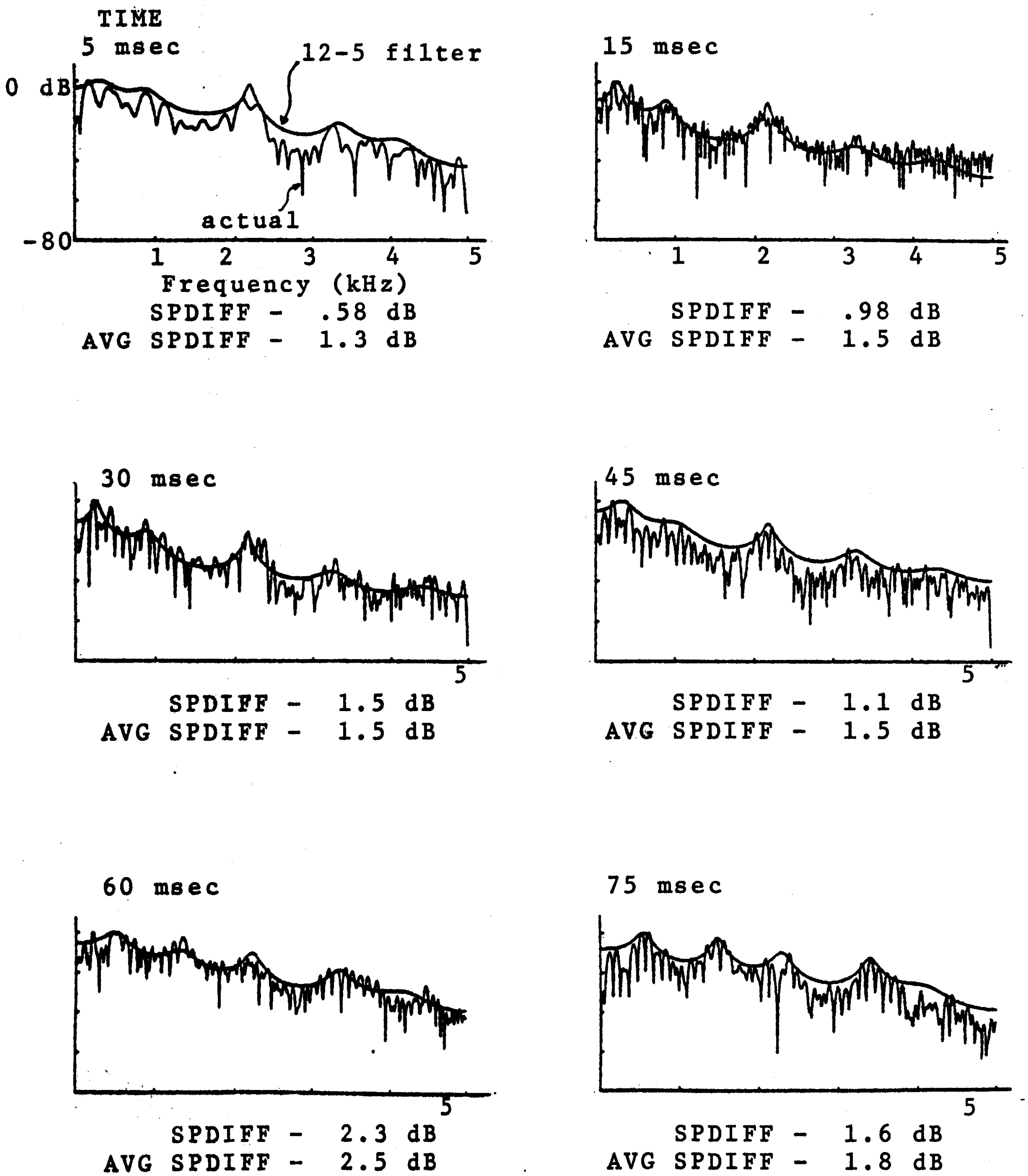

Figure 5.11 Comparison of Actual and 12-5 Covariance Power Filter Spectra (filter estimated for modified data)

are given in the figures. The equalization has resulted in the reduction of the average spectral difference to 1.5 dB for the time steps of 150, 300, and 450. These values are considerably lower than for the original 12-5 filter.

The only average spectral difference that was larger than 2.0 dB occurred for the interval around time step 600 (which is approximately the time of the abrupt change in system parameters). After time step 600, the average spectral differences were small.

This brief example of equalization would indicate that the method is better able to track the changing parameters throughout the entire interval if the signal is equalized. The best equalization would seem to be that based on equalizing the energy of the input impulses.

The next type of comparison that we have performed involves the impulse response for the original time-varying and time-invariant filters (estimated from the unequalized signal). The impulse response for both filters are shown in figures 5.12 and 5.13 for the various times indicated. The time-varying filter had an input train of impulses separated by 150 steps, so that each impulse occurred at the center of the corresponding LPC analysis interval. The impulse responses are almost identical after time step 600. However as the earlier analysis would indicate, there are significant differences for the times of 300 and 450. These figures give a visual indication of the severity of the spectral differences between the two methods.

As a final brief attempt to reproduce the original pre-emphasized

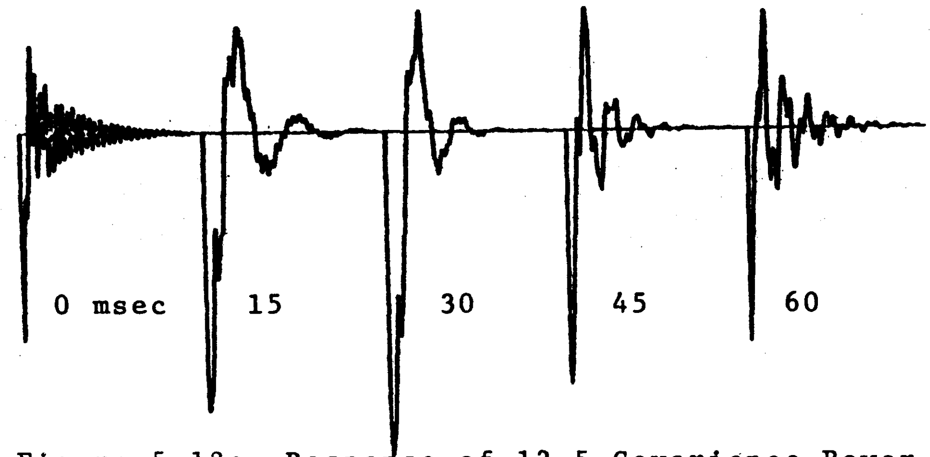

Figure 5.12a Response of 12-5 Covariance Power Filter to input train of impulses (the time of each input impulse is indicated below the corresponding oscillation)

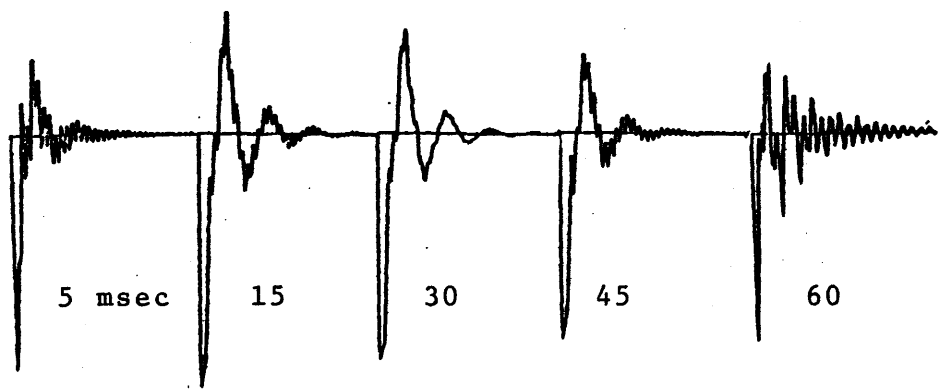

Figure 5.12b Impulse Responses of the 12-0 LPC Filters (center time of the analysis interval is indicated below the corresponding response)

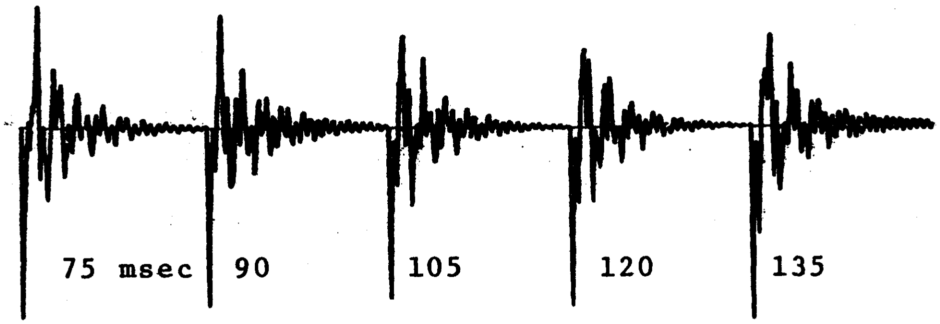

Figure 5.13a Response of 12-5 Covariance Power Filter to input train of impulses (continued)

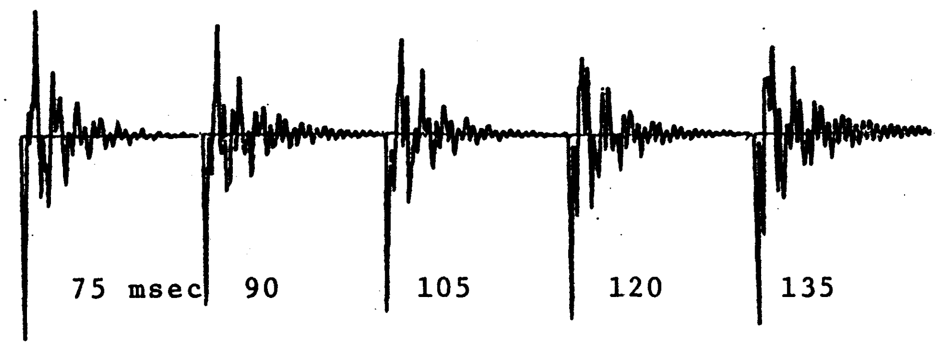

Figure 5.13b Impulse Responses of the 12-0 LPC Filters (continued)

signal, a 15-4 covariance power filter (no window) was estimated and used. (The 15-4 filter gave a better reproduction of the original than the 12-5 filter.) The input to the 15-4 power filter was a train of constant amplitude impulses separated by 100 data points, corresponding to a pitch period of 100 Hz. The reproduced signal is shown in figure 5.14.

The limitation of not having a time-varying gain estimation procedure is very evident in the reproduced signal. The magnitude of the signal is much too large at the beginning of the interval, and for the latter portion of the interval, the signal is too small. However the general characteristics of the original speech signal of figure 5.1b are there. The "low-pass" effect of the time-varying filter is evident in the time around 300-500.

In this chapter, we have examined the performance of time-varying LPC for one example of speech. This example was rather extreme in that there was a clear, significant, and relatively abrupt change in the shape of the waveform during the interval.

The "low-pass" effect of time-varying LPC was present but even so there was still fairly good agreement between the results for time-varying LPC and regular LPC. The significance of the low-pass effect was shown to be reduced by equalizing the signal or input impulse energy throughout the interval.

The attempt at reproducing the signal emphasized the need for a method to estimate the time-varying gain for the filter. However, this need might be eliminated by the use of signal equalization as

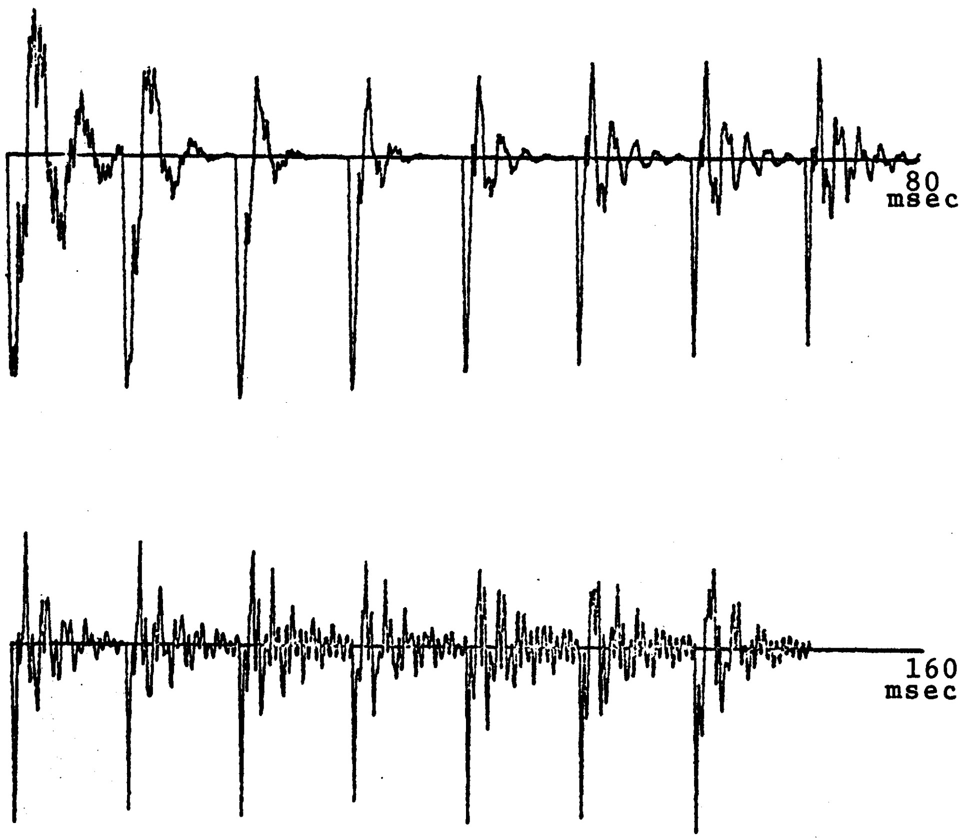

Figure 5.14 Reproduction of Original Signal Using 15-4 Covariance Power Filter.

mentioned above. If the equalization could be done in such a way so that the impulses driving the system could be thought of as approximately equal, then there would not be a time-varying gain. When attempting to reproduce the signal, the inverse of the signal equalization could be used.

CHAPTER VI

CONCLUSIONS

In this thesis, we have developed a method of time-varying linear prediction for the analysis of nonstationary speech signals. For this method, the coefficients of the speech production model were represented as linear combinations of a set of known time functions. In addition, an important contribution of this thesis is the investigation of methods for the evaluation of the performance of time-varying LPC. By using synthetic test cases, the general characteristics of time-varying linear prediction were determined. Time-varying LPC was shown to perform equally well when using either a power series or a Fourier series as the set of time functions. Since the time-varying method for the power series is computationally more efficient, the power series should be used as the set of time functions. It was demonstrated that time-varying LPC with the covariance method of error summation was better able to estimate the time-varying characteristics of the test cases than the autocorrelation method.

As discussed in the thesis, the autocorrelation method should not perform as well since it is based on an assumption that the speech waveform is stationary, which is not valid for this class of problems. In addition, we determined that the data should not be windowed because windowing degrades the accuracy of the estimation and also increases the likelihood that the estimated time-varying filter will have poles outside the unit circle.

We also demonstrated that the response of time-varying LPC to rapidly changing formant values is "low pass" in nature. Therefore this method is most effective in tracking slowly varying nonstationary speech characteristics, while for abrupt changes, it would provide a "smeared", less accurate estimate.

The performance of time-varying LPC for a speech example verified these characterizations of the method. It also demonstrated some of the limitations of the method. These limitations indicate the areas of future research for time-varying LPC.

The method does not perform as well for intervals of speech that contain an abrupt change in the system parameters. Therefore a method for detecting the abrupt changes needs to be developed. For this, the methods of failure detection [20] could possibly be used.

Another limitation of the method is that the resulting time-varying filter might be unacceptable because the "pole" trajectories may go outside the unit circle (as demonstrated by the filters estimated for the windowed speech data of Chapter V). The probability of this occuring is reduced if the data is not windowed; but even so, there is no guarantee that the time-varying filter will be stable. It may be possible to develop a time-varying estimation method that will necessarily result in a stable filter, however this has not been investigated in this thesis.

For the speech example, the time-varying filter "tracked" the parameters better during the high energy portions of the signal. This is a result of the least squares error technique of the method. One possible modification of the method to enable it to track the parameters equally well throughout the interval would be to have some form of automatic equalization of the signal. For this, the signal would be equalized so that it contains approximately equal energy throughout the interval. A simple way of implementing this would be to divide the interval into segments and estimate the energy in each segment (one estimate of the energy could be the $c_{00}(0,0)$ covariance element). The magnitude of each segment could be adjusted proportionally depending on whether its energy was above or below the average energy.

However, a more sophisticated technique might be necessary, because the equalization of the magnitude of the impulses driving the system is probably more important for the uniform tracking of the system parameters than the equalization of signal energy. Therefore the equalization should be also based on an estimate of the impulse magnitude.

Another serious limitation shown by the speech example is the lack of a time-varying gain estimate. Perhaps a method could be developed that would both equalize the signal in conjunction with providing a time-varying gain.

In spite of these limitations, the method of time-varying LPC seems promising. It can possibly reduce the total number of coefficients needed to model a segment of speech. It also provides a smoothed trajectory of the formants of the vocal tract. Time-varying LPC might also be used to provide higher quality speech reproduction than available with regular LPC if it is used over the same "quasi-stationary" intervals. This is because it can follow the small variation in the parameters that are present even in these "stationary" speech segments.

Additional research is needed to overcome the limitations of the method of time-varying LPC that has been developed in this thesis. Also, a more extensive evaluation of the method should be made by using it to analyze a wide variety of speech examples. Listening to speech reproduced by time-varying LPC should be an important part of future evaluation of the method.

REFERENCES

1. A.V. Oppenheim, ed., *Applications of Digital Signal Processing*, Chapter 3, Prentice-Hall, Inc., to appear in 1978.

2. G. Kopec, "Speech Analysis by Homomorphic Prediction", *S.M. Thesis*, Dept. of Electrical Engineering and Computer Science, Mass. Inst. of Tech., Cambridge, Mass., January 1976.

3. J.D. Markel and A.H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, New York, 1977.

4. B. Gold, "Note on Buzz-Hiss Detection", *J. Acoustic Soc. Amer.* Vol. 36, 1964, pp. 1659-1661.

5. J.D. Markel, "The SIFT Algorithm for Fundamental Frequency Estimation", *IEEE Trans. Audio and Electroacoustics*, Vol. AU-20, No. 5, December 1972, pp. 367-377.

6. J.I. Makhoul, "Linear Prediction: A Tutorial Review", *Proc. IEEE*, Vol. 32, April 1975, pp. 561-582.

7. R.W. Schafer and L.R. Rabiner, "Digital Representation of Speech Signals", *Proc. IEEE*, Vol. 32, April 1975, pp. 662-677.

8. A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

9. A. Gelb, ed., *Applied Optimal Estimation*, M.I.T. Press, Cambridge, Mass., 1974.

10. L.A. Liporace, "Linear Estimation of Nonstationary Signals", *J. Acoust. Soc. Amer.*, Vol. 58, No. 6, pp. 1268-1295, December 1975.

11. R.A. Wiggins and E.A. Robinson, "Recursive Solutions to the Multichannel Filtering Problem", *J. Geophys. Res.*, Vol. 70, 1965, pp. 1885-1891.

12. E.A. Flinn, "A Theoretical Discussion of Optimum Multichannel Filter Design", Rept. AD 845643, prepared for Air Force Technical Applications Center, December 10, 1968.

13. J. Rissanen, "Algorithms for Triangular Decompositions of Block Hankel and Toeplitz Matrices with Applications to Factoring Positive Matrix Polynomials", *Mathematics of Computation*, Vol. 27, 1973, pp. 147-154.

14. J.J. Cornyn, Jr., "Direct Methods for Solving Systems of Linear Equations Involving Toeplitz or Hankel Matrices", Rept. AD/A-002 931, prepared for Office of Naval Research, October 1974; also S.M. Thesis, University of Maryland, College Park, Md.

15. M. Morf, T. Kailath, and B. Dickinson, "Notes on a Fast Algorithm for Solving Atal's Normal Equations", unpublished paper presented at Arden House Workshop, February, 1976.

16. G.C.M. Fant, Acoustic Theory of Speech Production, Mouton and Co., 's-Gravenhage, The Netherlands, 1960.

17. A.H. Gray, Jr., and J.D. Markel, "Distance Measures for Speech Processing", IEEE Trans. Acous., Speech and Sig. Proc., Vol. 24, No. 5, October 1976, pp. 380-391.

18. J. Turner and B. Dickinson, "Linear Prediction Applied to Time-Varying All Pole Signals", in Proc. 1977 IEEE Int. Conf. on Acous., Speech and Sig. Proc., pp. 750-753.

19. S.D. Conte and C. deBoor, Elementary Numerical Analysis, McGraw-Hill, Inc., New York, 1972.

20. A.S. Willsky, "A Survey of Design Methods for Failure Detection in Dynamic Systems", Automatica, Vol. 12, November 1976, pp. 601-611.

21. M. Morf, B. Dickinson, T. Kailath, and A. Vieira, "Efficient Solutions of Covariance Equations for Linear Prediction," IEEE Trans. Acous., Speech and Sig. Proc., to appear 1977.