Signal Processing 5 (1983) 267–285
North-Holland Publishing Company

TIME-VARYING PARAMETRIC MODELING OF SPEECH*

Mark G. HALL

Naval Surface Weapons Center, DK-51, Dahlgren, VA 22448, USA

Alan V. OPPENHEIM

Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Room 36-615, Cambridge, MA 02139, USA

Alan S. WILLSKY

Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science and Laboratory for Information and Decision Systems, Room 35-233, Cambridge, MA 02139, USA

Received 16 March 1982
Revised 1 November 1982

Abstract. For linear predictive coding (LPC) of speech, the speech waveform is modeled 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 paper 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.

Zusammenfassung. Bei der Linearen Prädiktion (LPC) von Sprache wird die Sprachzeitfunktion modellhaft als Ausgangssignal eines Allpole-Filters aufgefaßt. Die Zeitfunktion wird dabei in zahlreiche kurze Intervalle von 10 bis 30 ms Dauer unterteilt, in denen das Signal als stationär betrachtet werden kann. Für jedes Intervall werden die konstanten Koeffizienten des Allpole-Filters durch Lineare Prädiktion ermittelt, wobei ein quadratisches Prädiktions-Fehlermaß minimisiert wird. In der vorliegenden Arbeit wird eine Modifikation des LPC-Verfahrens vorgestellt - das sog. Zeitvariante LPC-Verfahren - mit dessen Hilfe es möglich ist, nicht-stationäre Sprachsignale zu analysieren. Bei diesem Verfahren dürfen die Koeffizienten des Allpole-Filters variant sein unter der Voraussetzung, daß sie sich als eine Linearkombination eines Satzes bekannter Zeitfunktionen darstellen lassen. Die Koeffizienten der Linearkombination von Funktionen erhält man mit Hilfe der gleichen Technik des kleinsten Fehlerquadrats, wie sie auch beim LPC-Verfahren verwendet wird. Es werden Methoden zur Messung und Überprüfung der Leistungsfähigkeit des Zeitvarianten LPC-Verfahrens entwickelt und Ergebnisse von Verfahren der Zeitvarianten LPC-Analyse sowohl von synthetisch erzeugter als auch von echter Sprache mitgeteilt.

Résumé. Pour le codage prédictif (LPC) de la parole, on modélise l'onde de parole comme la sortie d'un filtre tout-pole. Cette onde est divisée en de nombreux intervalles de courte durée (10–30 msec), pendant lesquels le signal de parole est considéré comme stationnaire. Pour chaque intervalle, les coefficients constants du filtre tout-pole sont estimés en prédiction linéaire par minimisation d'un critère quadratique de l'erreur de prédiction. Cet article étudie une modification de la prédiction linéaire, appelée prédiction linéaire variable dans le temps, qui peut être utilisée por analyser des signaux de

* This work was conducted in part at the M.I.T. Research Laboratory of Electronics with partial support provided by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506, and in part at the M.I.T. Laboratory for Information and Decision Systems with partial support provided by NASA Ames Research Center under Grant NGL-22-009-124.

parole non stationnaires. Dans cette méthode, chaque coefficient du filtre tout-pole est autorisé à varier dans le temps en considérant qu'il est combinaison linéaire d'un ensemble donné de fonctions du temps. Les coefficients de la combinaison linéaire de fonctions sont obtenus de la même façon qu'en LPC, avec la technique des moindres carrés. Des méthodes sont développées pour mesurer et évaluer les performances de la prédiction linéaire variable dans le temps, et les résultats de cette LPC variable sont présentés tant pour des signaux de parole synthéetiques que pour des signaux de parole réels.

Keywords. Autoregressive models, nonstationary signals, parameter identification, speech.

1. Introduction

Parametric analysis and modeling of signals using an autoregressive model with constant coefficients has found application in a variety of contexts including speech and seismic signal processing, spectral estimation, process control and others. In many cases, the signal to be modeled is time-varying. However, if the time variation is relatively slow, it is nevertheless reasonable to apply a constant model on a short-time basis, updating the coefficients as the analysis proceeds through the data [1, 2].

In this paper, we consider autoregressive signal modeling in which the coefficients are time-varying. In our method, each coefficient in the model is allowed to change in time by assuming it is a linear combination of some set of known time functions. Thus each autoregressive coefficient is itself specified by a set of parameters, the coefficients in the linear combination. Using the same least-squares error technique as used for modeling with constant coefficients (specifically LPC as outlined in Section 2), the parameters in the linear combinations for all of the autoregressive coefficients can be found by solving a set of linear equations. Therefore the determination of the model parameters for time-varying LPC is similar to that for traditional LPC, but there is a large number of coefficients that must be obtained for a given order model.

There are several potential advantages to time-varying LPC. In some cases the system model may be more realistic since it allows for the continuously changing behavior of the signal. This should lead to increased accuracy in signal representation. In addition, the method may be more efficient since the inclusion of time variations in the model should allow analysis over longer data windows. Therefore, even though time-varying LPC involves a larger number of coefficients than traditional LPC, it will divide the signal into fewer segments. This could result in a possible reduction of the total number of parameters needed to accurately model a segment of data 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 time-varying signal modeling methods in general and time-varying LPC in particular. One of the goals of this work has been to explore methods for understanding the behavior of time-varying models and for evaluating their performance. Several such techniques are used in this paper and should be of some independent interest.

In the next section we formulate the problem of time-varying LPC and derive the basic equations. Computational aspects of this approach are addressed in Section 3. In Section 4, we present and discuss methods for evaluating time-varying linear prediction and we apply these methods to some experimental results for synthetic speech waveforms. In Section 5, we compare the results of time-varying LPC and time-invariant LPC analysis for an actual speech waveform. The results of this analysis are of interest since a longer analysis window was used for time-varying LPC than for time-invariant LPC.

2. Time-varying linear prediction

For all-pole signal modeling, the signal $s(n)$ at time n is modeled as a linear combination of the past p samples and the input $u(n)$, i.e.,

$$s(n) = -\sum_{i=1}^{p} a_i s(n-i) + Gu(n). \qquad (2.1)$$

The method of linear prediction (or linear predictive coding, LPC) is typically used to estimate the coefficients and the gain factor [1, 2]. In this approach it is assumed that the signal is stationary over the time interval of interest and therefore the coefficients given in the model of (2.1) are constants. For speech, for example, this is a reasonable approximation over short intervals (10–30 msec).

For the method of time-varying linear prediction, the prediction coefficients are allowed to change with time, so that (2.1) becomes

$$s(n) = -\sum_{i=1}^{p} a_i(n) s(n-i) + Gu(n). \qquad (2.2)$$

With this model, the signal is not assumed to be stationary and therefore the time-varying nature of the coefficient $a_i(n)$ must be specified. We have chosen to model these coefficients as linear combinations of some known functions of time $u_k(n)$:

$$a_i(n) = \sum_{k=0}^{q} a_{ik} u_k(n). \qquad (2.3)$$

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 $u_0(n) = 1$.

By limiting our attention to such a model, we are clearly constraining the possible types of time variations that can be modeled. However, if we allowed arbitrary variations in the coefficients, we would have as many degrees of freedom in the parametric model as in the original data, thus achieving no data compression or insight into the structure of the signal. Thus constraints on the nature of the time variations are essential. However, by judicious choice of the basis functions $u_k(n)$ we can accurately approximate a wide variety of coefficient time variations. Possible sets of functions that could be used include powers of time

$$u_k(n) = n^k \qquad (2.4)$$

or trigonometric functions as in a Fourier series

$$\begin{aligned} u_k(n) &= \cos(k\omega n), \quad k \text{ even}, \\ u_k(n) &= \sin(k\omega n), \quad k \text{ odd} \end{aligned} \qquad (2.5)$$

where ω is a constant dependent upon the length of the speech data. In particular, we have chosen $\omega = \pi/N$, where N is the total number of data points in the speech data. The reason for this choice is that *any* time varying signal $a(n)$ can be represented *exactly* as in (2.3) if we let $q \to \infty$ and use the $u_k(n)$ in (2.5) with this choice of ω. Note that a choice of ω equal to $2\pi/N$ or larger would force $a_i(n)$ in (2.3) to be periodic with period less than N (for example $\omega = 2\pi/N$ would lead to the condition $a_i(N) = a_i(0)$). Any choice of $\omega < 2\pi/N$ avoids this constraint, and our particular choice leads to some computational simplifications. Liporace [3] seems to have been the first to have formulated the problem as in (2.3). His analysis used the power series of the form of (2.4) for the set of functions. See also [10] which presents a general framework for estimating nonstationary ARMA models.

From (2.2) and (2.3), the predictor equation is given as

$$\hat{s}(n) = -\sum_{i=1}^{p} \left(\sum_{k=0}^{q} a_{ik} u_k(n) \right) s(n-i) \qquad (2.6)$$

and the prediction error is

$$e(n) = s(n) - \hat{s}(n). \qquad (2.7)$$

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

$$E = \sum_n e^2(n) = \sum_n \left(s(n) + \sum_{i=1}^{p} \sum_{k=0}^{q} a_{ik} u_k(n) s(n-i) \right)^2. \qquad (2.8)$$

Minimizing the error with respect to each coefficient and defining the generalized correlation function

$$c_{kl}(i,j)=\sum_n u_k(n)u_l(n)s(n-i)s(n-j), \qquad (2.9)$$

the coefficients are specified by the equation

$$\sum_{i=1}^{p}\sum_{k=0}^{q} a_{ik}c_{kl}(i,j)=-c_{0l}(0,j)$$

$$1\leq j\leq p, \quad 0\leq l\leq q. \qquad (2.10)$$

For the correlation function $c_{kl}(i,j)$, 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 correlation function $c_{00}(i,j)$ is the same as the LPC correlation function.

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 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. For the covariance method, the sum over n goes from p to $N-1$, and (2.10) can be expressed in matrix form by defining the vectors

$$\boldsymbol{a}_i^{\mathrm{T}}=[a_{1i}, a_{2i}, a_{3i}, \ldots, a_{pi}], \quad 0\leq i\leq q \qquad (2.11)$$

and

$$\boldsymbol{\psi}_i^{\mathrm{T}}=[c_{0i}(0,1), c_{0i}(0,2), \ldots, c_{0i}(0,p)],$$

$$0\leq i\leq q, \qquad (2.12)$$

and the matrix

$$\Phi_{kl}=\begin{bmatrix} c_{kl}(1,1) & c_{kl}(1,2) & \cdots & c_{kl}(1,p)\\ c_{kl}(2,1) & c_{kl}(2,2) & \cdots & c_{kl}(2,p)\\ \vdots & \vdots & & \vdots\\ c_{kl}(p,1) & c_{kl}(p,2) & \cdots & c_{kl}(p,p)\end{bmatrix}$$

$$0\leq k\leq q, \quad 0\leq l\leq q. \qquad (2.13)$$

From (2.9) it is clear that $\Phi_{kl}=\Phi_{lk}=\Phi_{kl}^{\mathrm{T}}$ so that (2.10) becomes

$$\begin{bmatrix} \Phi_{00} & \Phi_{01} & \cdots & \Phi_{0q}\\ \Phi_{10} & \Phi_{11} & \cdots & \Phi_{1q}\\ \vdots & \vdots & & \vdots\\ \Phi_{q0} & \Phi_{q1} & \cdots & \Phi_{qq}\end{bmatrix}\begin{bmatrix}\boldsymbol{a}_0\\ \boldsymbol{a}_1\\ \vdots\\ \boldsymbol{a}_q\end{bmatrix}=-\begin{bmatrix}\boldsymbol{\psi}_0\\ \boldsymbol{\psi}_1\\ \vdots\\ \boldsymbol{\psi}_q\end{bmatrix} \qquad (2.14)$$

or

$$\Phi\boldsymbol{A}=-\boldsymbol{\Psi}. \qquad (2.15)$$

Because $\Phi_{kl}=\Phi_{lk}=\Phi_{kl}^{\mathrm{T}}$, Φ is a $(q+1)\times(q+1)$ block symmetric matrix with $(p\times p)$ symmetric blocks. Equation (2.15) can alternately be expressed so that Φ is a $(p\times p)$ block symmetric matrix with $(q+1)\times(q+1)$ symmetric blocks (see [4]).

A similar, but not identical, set of equations, analogous to the autocorrelation method in the time-invariant case, can be formulated by windowing the data and minimizing the error over an infinite time interval. In this formulation, in order for the matrix Φ in (2.15) to be expressed as a block Toeplitz matrix, (2.3) is modified to

$$a_i(n)=\sum_{k=0}^{q} a_{ik}u_k(n-i), \quad 1\leq i\leq p. \qquad (2.16)$$

The minimization of the total error again results in eq. (2.10) with the autocorrelation coefficients defined as

$$\begin{aligned} c_{kl}(i,j) &= \sum_{n=-\infty}^{\infty} u_k(n-i)u_l(n-j)s(n-i)s(n-j)\\ &= \sum_{n=-\infty}^{\infty} u_k(n)u_l(n+i-j)s(n)s(n+i-j)\\ &\triangleq r_{kl}(i-j). \end{aligned} \qquad (2.17)$$

With this definition of the autocorrelation coefficients, the matrix Φ in (2.15) is symmetric and block Toeplitz. Note that (2.16) does not represent a substantive change and is merely a rearrangement that leads to a particularly nice structure for Φ.

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 signal 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 distortions of the LPC coefficients due to the discontinuities in the data at the ends of the interval evidenced in the time-invariant case apply also 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.

Windowing of the signal is a usual practice for the LPC autocorrelation 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 tends to 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 the possible adverse effects on the estimation of the prediction coefficients due to an additional time variation caused by windowing the data, the use of a window does not seem beneficial for either the autocorrelation or covariance method. However for the autocorrelation method, there are distortions in the estimates caused by the end effects when the data is not windowed. Therefore the autocorrelation method seems to have more disadvantages than the covariance method. This conclusion is supported by the fact that for time-invariant LPC, there is a basic assumption for the autocorrelation method that the signal is stationary. The use of the autocorrelation method for time-varying LPC would imply a contradiction of this basic assumption since the signal is definitely not stationary. Nevertheless, we have implemented this method in order to observe its behavior.

To completely represent the signal by the model of (2.2), the prediction coefficients $a_i(n)$, the gain factor G, and the input $u(n)$ must all be estimated. The analysis of this paper is concerned primarily with the estimation of the time-varying prediction coefficients, while the problem of estimating the gain and the input for the time-varying model has not been addressed. However, the effect of the time-varying gain on the estimation of the prediction coefficients will be discussed in Sections 5 and 6.

3. Computational aspects of time-varying linear prediction

For the time-varying linear prediction method outlined in Section 2, the predictor coefficients (a_{ik}, $1 \leq i \leq p$, $0 \leq k \leq q$) are obtained by solving a set of linear equations given by (2.10). Because the number of coefficients increases linearly with the number, $(q+1)$, of terms in the series expansion, there is a significant increase in the amount of computation for time-varying LPC as compared with traditional LPC (where $q=0$). Some of the techniques discussed in reference [4] can be used to make the coefficient determination efficient. This section is an initial discussion of the computational aspects, and no detailed algorithm has been developed.

There are four possible techniques (covariance-power, covariance Fourier, autocorrelation-power, and autocorrelation-Fourier) that can be used for time-varying LPC since there are two methods of summation (covariance or autocorrelation) for (2.9) and two possible sets of basis functions (power or Fourier series) that we have used for the prediction coefficients. For any of these techniques, the computations needed for the determination of the coefficients can be divided into two categories. Most of the computational effort is involved with calculating the elements of Φ and ψ. Then once these elements have been determined, the normal equations of (2.15) must be solved.

There are $p^2(q+1)^2$ elements in the matrix Φ and $p(q+1)$ elements in the vector ψ that need to be calculated. However, Φ is symmetric for both the covariance and autocorrelation methods, which reduces the number of distinct matrix elements to be computed to $\frac{1}{2}p(q+1)[p(q+1)+1]$. But because Φ may have additional symmetry, this number can be reduced further.

For the covariance method (either for power or Fourier series), the matrix elements of (2.9) have the additional symmetry that

$$c_{kl}(i,j)=c_{lk}(i,j)=c_{kl}(j,i)=c_{kl}(j,i).$$

Therefore the matrix Φ can be expressed as a block symmetric matrix with each block being a symmetric matrix. Because of this symmetry only $\frac{1}{4}p(p+1)(q+1)(q+2)$ distinct elements for the matrix Φ need to be calculated. This number can be reduced even more for the covariance power series method, because for $k+l=m$, $u_k(n)u_l(k)=u_m(k)=n^m$, and thus

$$c_{m0}(i,j)=c_{kl}(i,j), \tag{3.1}$$

so that only the elements $c_{k0}(i,j)$, $0\le k\le 2q$, need to be computed. Because of this symmetry Φ can be expressed as a block Hankel matrix (where all the block matrices along the secondary diagonal, northeast to southwest, are equal) for the covariance power series method. With the additional symmetry, only $[\frac{1}{2}p(p+1)](2q+1)$ elements must be computed for the Φ matrix of the covariance power series method.

For the autocorrelation method (for either the power or Fourier series) it can be shown that

$$r_{kl}(i-j)=r_{lk}(j-i). \tag{3.2}$$

With this symmetry, the number of distinct elements of Φ for the autocorrelation methods is $p(q^2+2q+1)-\frac{1}{2}(q+1)q$.

These results show that, in general, there is a slightly smaller number of unique elements in the matrix Φ for the autocorrelation methods than for the covariance power series method. The covariance Fourier series method has significantly more elements that must be calculated than the other methods. For any of these methods, only some of the elements have to be calculated using the summation given by (2.9) or (2.17), for many elements can be calculated recursively from previously computed elements. As an example, for the covariance power series method, it can be shown (see [4]) that

$$\begin{aligned} c_{mo}(i,j) &= \sum_{r=0}^{n} \frac{m'}{(m-r)!r!} c_{m-r,0}(i-1,j-1) \\ &\quad + p^m s(p-i)s(p-j) \\ &\quad - N^m s(N-i)s(N-j). \end{aligned} \tag{3.3}$$

Recursions can be developed for the covariance Fourier series method and the autocorrelation methods, however they cannot be expressed as compactly as the recursion of (3.3). Since the covariance Fourier series or autocorrelation elements do not have the symmetry of the covariance power series elements as shown in (3.1) more elements must be computed using the summations of (2.9) for these methods. It should be noted that the determination of the individual matrix elements is faster for the power series methods than for the Fourier series methods because no trigonometric functions need to be evaluated.

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 $k = l = 0$, 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.

Once the elements of the matrix Φ have been calculated, the set of equations must be solved to determine the coefficients. Liporace [3] has developed an efficient algorithm to solve the equations for the covariance method where Φ is a block symmetric matrix with symmetric blocks. The covariance method using the power series has the additional advantage that Φ can be expressed as a block Hankel matrix for which there is an efficient solution [5, 6]. For the autocorrelation method, Φ is a block Toeplitz matrix and there is an algorithm given in reference [7] for solving the equations. This method is an extension of Levinson's recursive algorithm to the multichannel filtering problem. Thus, the autocorrelation power series method yields the set of equations that can be solved most efficiently, as the matrix Φ can be expressed either as a block Hankel matrix with Toeplitz blocks or a block Toeplitz matrix with Hankel blocks.

4. Experimental results for synthetic data

For the evaluation of time-varying linear prediction, one method used was 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 excited by a periodic impulse train 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 Section 2), and between windowing or not windowing the signal.

An example of one of the test cases is shown in Fig. 1(a). The signal was generated by a 6-2 power series filter.[1] The sampling rate for this example was 10 kHz and the period of the excitation impulse train was 100 samples, corresponding to a fundamental frequency of 100 Hz.

In order to determine how well the estimated time-varying filters could reproduce the original signal, the impulse train used to generate the data was passed through the estimated filters. The response of the estimated 6-2 covariance power filter (with or without windowing) was virtually identical to the original and therefore is not shown. Similarly, the 6-2 covariance Fourier filter (no windowing) was also virtually identical to the original. The 6-2 autocorrelation power filter response (no windowing) is shown in Fig. 1(b) and the 6-2 autocorrelation power filter response (windowing) is shown in Fig. 1(c). The response of a 6-4 autocorrelation Fourier filter (with or without windowing) was essentially identical to that of the 6-2 autocorrelation power filter (Fig. 1(c)). It can be seen that the major differences between the original data and the responses of the filters determined using the autocorrelation method occur at both ends of the interval. The response of the filter determined without windowing the data does not match the original data as well as the response estimated with windowing.

As another method of evaluating 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

[1] A 6-2 power series filter has 6 poles, ($p = 6$), with each coefficient being a quadratic power series ($q = 2$).

Fig. 1(a). Synthetic speech example generated by 6-2 power series filter.

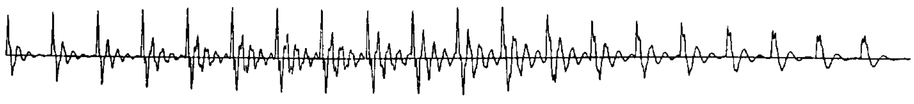


Fig. 1(b). Response of 6-2 autocorrelation power filter (without windowing the original data).

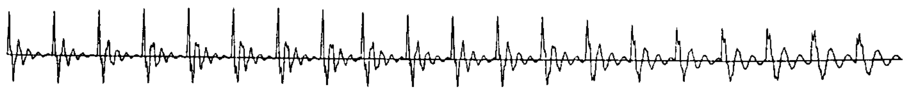


Fig. 1(c). Response of 6-2 autocorrelation power filter (windowing the original data).

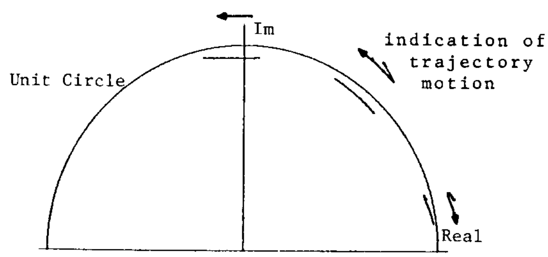
interval $[0, N-1]$), where $p(z, n)$ is defined as

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

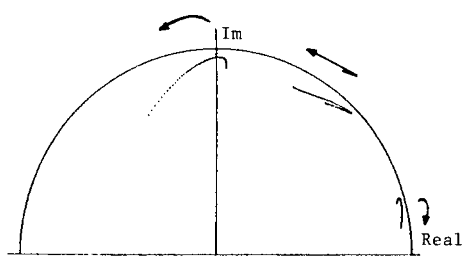
Strictly speaking, it is only correct to talk about poles for time-invariant systems. However, for systems that vary slowly with time, it may be useful conceptually to think of them as possessing time-varying poles. When these 'poles' change slowly in time, one should be able to deduce some qualitative aspects of the system behavior by observing the pole trajectories. Thus one possible measure of performance is the ability of our parameter estimation system to track these poles.

Two filters with different pole trajectories are necessarily significantly different in impulse response or general characteristics. However, 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.

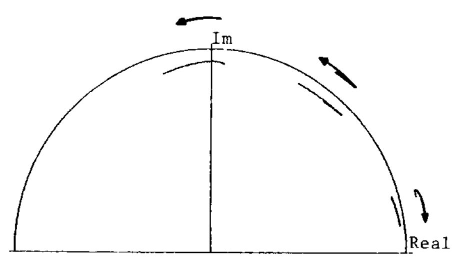
Fig. 2 shows the pole trajectories of the filters using the estimated coefficients with the power series method. The graphs plot the real part of each pole on the ordinate and the imaginary part on the abscissa. The location of each pole of the filter is plotted every 25 msec of the analysis interval. The unit circle is also shown on the graphs for comparison purposes.



(a) 6-2 covariance power filter (without window)



(b) 6-2 autocorrelation power filter (without window)



(c) 6-2 autocorrelation power filter (with window)

Fig. 2. Pole trajectories for power series.

Fig. 2(a) shows the pole trajectories for the 6-2 filter estimated by using the covariance power series method with no windowing. Since these trajectories closely matched the pole trajectories of the generating filter and of the estimated 6-2 covariance power series filter using a Hamming window, these other trajectories are not shown. The only minor differences between the windowed and non-windowed cases occur at each end of the trajectory, where the effect of the window is the most significant. Figs. 2(b) and 2(c) 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 Fig. 2(a). This would seem to verify the suggestion in Section 2, that the autocorrelation method attempts to minimize (unrealistically) the error at the extreme ends of the interval, and consequently, there might be some distortion in the coefficients at the ends.

Fig. 2(c) 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 Fig. 2(b). In fact, these trajectories compare favorably with those of Fig. 2(a). The only major differences are those of the third pole.

The figures for the pole trajectories of the 6-4 covariance or autocorrelation Fourier series method are not shown because they are almost identical with the respective power series trajectories.

There were many conclusions to be drawn from these examples. The differences between using a power series or a Fourier series for the analysis seem 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, but this was not a general rule.

The use of a window had only a slight effect 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 method without a window.

These results can be explained, however, by the fact that the test cases were generated by a system whose actual 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, for actual signals not generated by a system exactly of the form assumed in time-varying LPC, 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. Thus, it does not seem that windowing is generally a good practice. In Section 5, the effect of windowing actual nonstationary speech on the analysis results will be shown.

All of our analysis and experience, both with synthetic and real data, indicate that the covariance method without windowing should be used. Since the results seem to be similar for either the power or Fourier series, the power series form seems preferable because of its computational advantage over the Fourier series method, as discussed in Section 3.

The second set of synthetic data experiments involved the response of time-varying LPC to step changes in the center frequency of the poles of the system generating the data. Abrupt signal changes such as this represent an extreme form of nonstationarity, and in order to increase our understanding of time-varying LPC it is of interest to see how it performs under such conditions. In addition, many signal processing problems involve data containing such changes. For example, in the speech context, consonants and plosives might be modeled in this manner. Thus these experiments have both conceptual and practical motivations.

The study was carried out using a four-pole system for the generation of the synthetic signal. The center frequency of two poles changes discontinuously. The 4-3 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 Fig. 3. Note that the trajectories of the poles for the time-varying linear prediction method resemble the step response of a non-causal low-pass filter. The fact that the response is anticipative is not surprising, since the entire data interval is used to estimate the coefficients. In addition, our results indicate that the time-varying LPC system response is approximately homogeneous in that the pole angle trajectory for a given center frequency change is proportional to the size of the step change and is approximately additive in that the response to two different jumps in one interval is approximately the same as the sum of the responses to each jump taken separately in the same interval. Thus, the method can be thought of as acting like a linear lowpass filter in response to changes in the location of the poles. An estimate of the frequency response of the method's lowpass action was obtained from the computed step responses and is shown in Fig. 4 for the 4-3 and 4-5 covariance power filter. The 4-5 method has a broader 'frequency response'. Since we can fit higher order coefficient fluctuations using the higher order series expansion, the greater sensitivity of the 4-5 method to high frequency changes is not surprising.

The pole trajectories for the 4-3 and 4-5 covariance power methods are compared with the response for the traditional LPC covariance method (the 4-0 covariance method) in Fig. 5. Unlike the time-varying methods which used the entire 60 msec. data interval, a data interval of 15 msec. was used for the time-invariant LPC

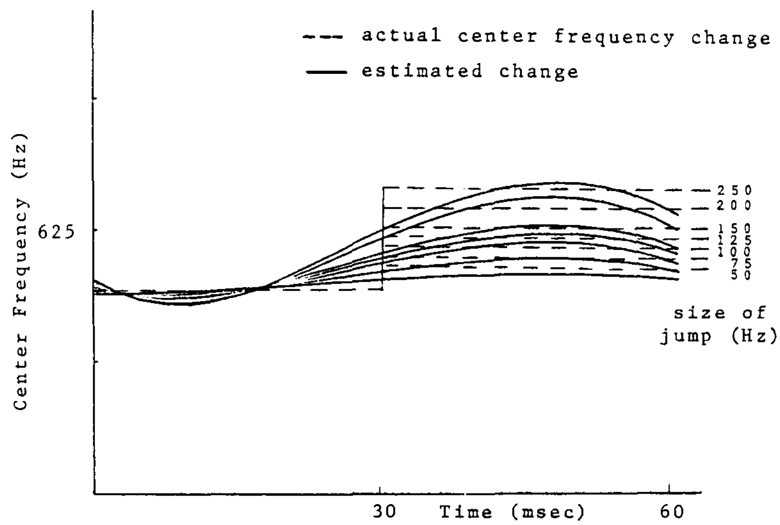


Fig. 3. Center frequency trajectories for 4-3 covariance power filter.

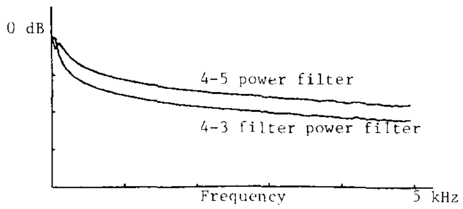


Fig. 4. Comparison of smoothed unit pulse frequency responses.

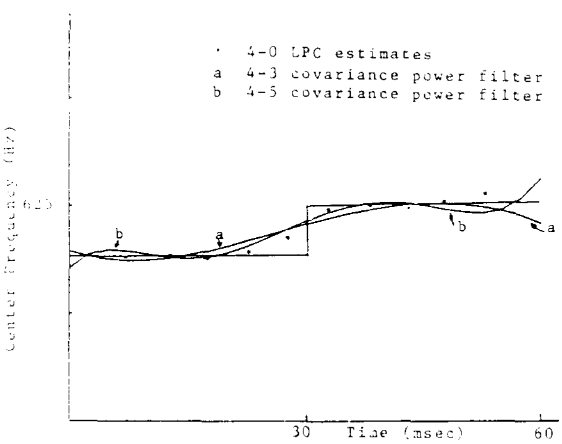


Fig. 5. Center frequency trajectories for 150 Hz jump.

method. The starting location of the analysis interval was shifted by 5 msec. for each successive LPC analysis, so that there was some overlap of the data on each interval. Over the full 60 msec., this means that 10 separate LPC analyses were performed. The center frequency of the pole for each interval is plotted at the center of the time interval. No windowing was used.

From Fig. 5 we see 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. The interpretation of this result, however, requires some thought. Note that in this analysis we are effectively viewing time-invariant LPC as a *time-varying* identification method by examining its behavior over successive analysis intervals. In this sense, it is not surprising that time-invariant LPC has a fast response to the abrupt change, since it has less memory than the time-varying methods, and it changes discontinuously as the parameters are updated at successive intervals. In addition, note that the 4-0 covariance method, that is, traditional LPC, requires a total of 40 coefficients over the entire 60 msec. segment (4 coefficients for each of the 10 analysis intervals). In contrast, the 4-3 covariance method requires 16 coefficients, while the 4-5 method needs 24 coefficients.

Finally, test cases were also run to evaluate the ability of the method to track slowly varying changes. Specifically, the first pole was varied linearly in frequency over the interval, while the second pole was held constant. Examples of the change in frequency for the first pole are shown in Fig. 6. The changes in the center frequency of the first pole for the estimated 4-3 covariance power series filter are also shown.

It can be seen from the figures 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 indicated that the method acted as a lowpass filter. Evidently, the higher slope changes are beyond the cutoff frequency of the method, yielding the same estimated pole trajectory as for an abrupt step change.

5. Experimental results for time-varying analysis of speech

In this section, we give an example of the application of time-varying LPC to a nonstationary speech waveform, shown in Fig. 7. In order to estimate the spectral properties of the vocal tract, the waveform was pre-emphasized by a simple one zero filter in the form of $1-\mu z^{-1}$ ($\mu = 0.95$) to remove the glottal effects. Several different methods for evaluating the performance were used. The pole trajectories of time-varying LPC were compared with the poles of the time-invariant filters estimated by regular LPC. The log

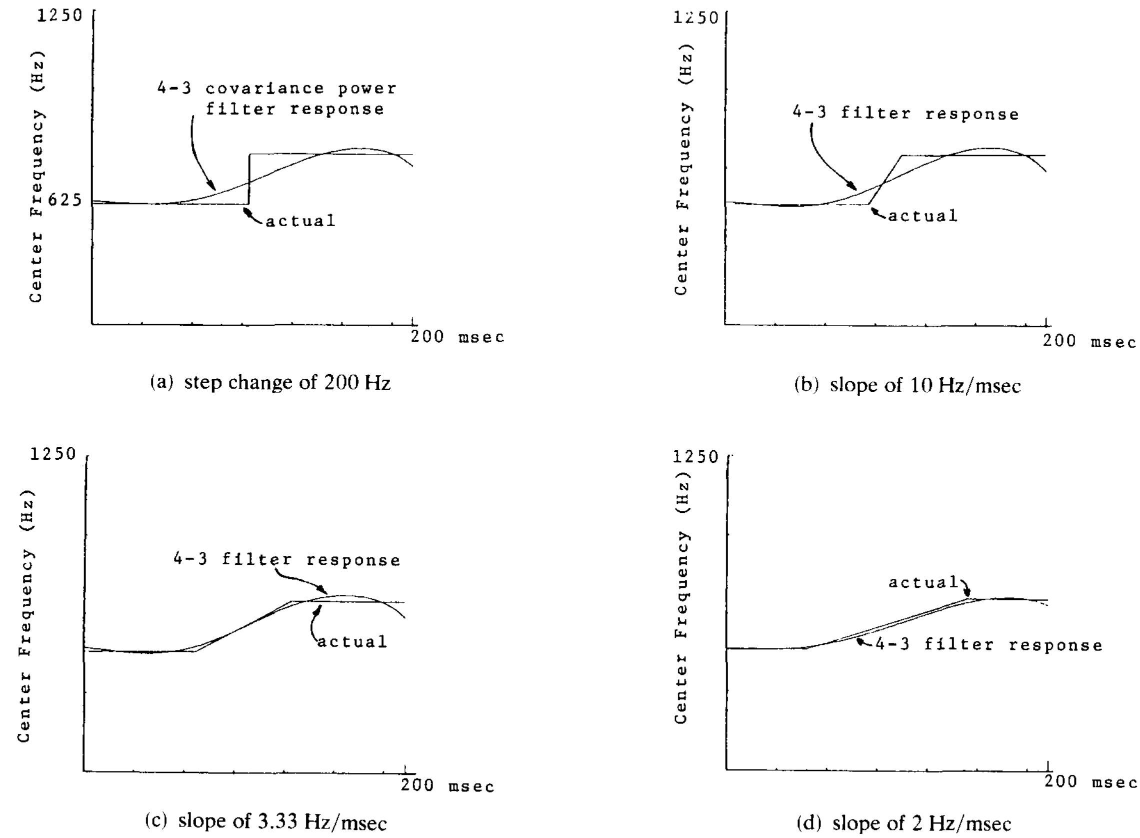


Fig. 6. Center frequency trajectories.

spectrum of each time-invariant LPC filter was also 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 Gray and Markel [8] and Turner and Dickinson [9] was used. In addition, the impulse response of both regular and time-varying LPC were compared with the original speech data. The time-varying model that was used was a 12-5 power series filter, and the analysis was performed on an interval of length 150 msec.

For regular LPC, a 12-pole filter was used and the length of each analysis interval was 20 msec. The center of the interval was shifted by 15 msec for each successive LPC analysis, resulting in some overlap of the data contained in each interval.

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. For any of these methods, there was no attempt to estimate the gain for this speech signal, although it does seem evident from the speech signal that the gain was changing with time.

The pole trajectories for the covariance power series method both with and without windowing the data are shown in Fig. 8. These results dramatically illustrate the effect of windowing, since for the windowed data some of the poles of the model leave the interior of the unit circle, while they do not when the data is not windowed. For a time-invariant filter, this would mean that the filter was

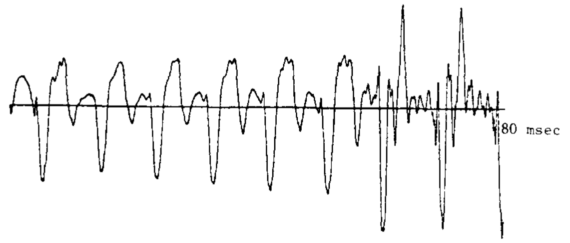


Fig. 7(a). Nonstationary speech waveform.

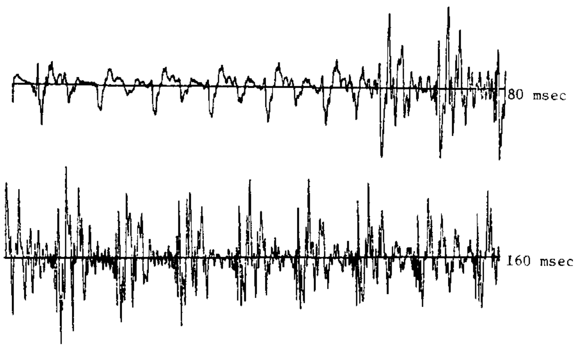


Fig. 7(b). Pre-emphasized version of speech waveform of Fig. 7(a).

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 Fig. 9. Again, the autocorrelation filter for the windowed data has poles outside the unit circle. The results of the autocorrelation method (without windowing) agree favorably with those of the covariance method. The most significant differences occur at

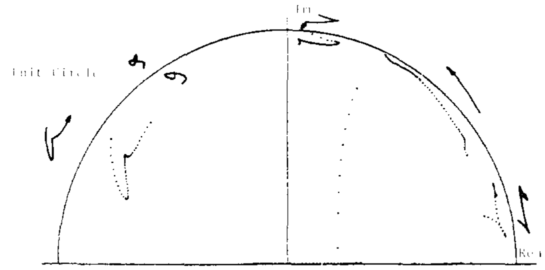


Fig. 8(a). Pole trajectories for 12-5 covariance power filter (data not windowed).

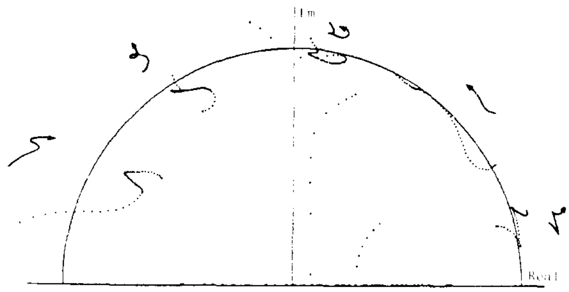


Fig. 8(b). Pole trajectories for 12-5 covariance power filter (data windowed).

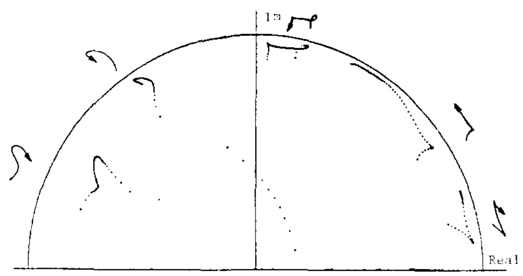


Fig. 9(a). Pole trajectories for 12-5 correlation power filter (data not windowed).

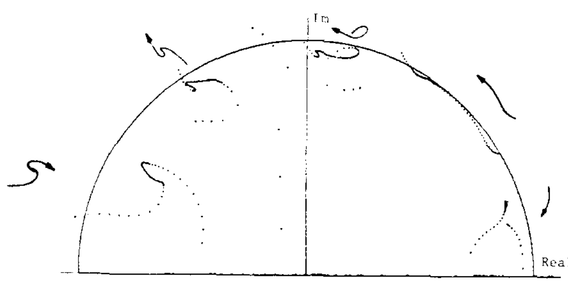


Fig. 9(b). Pole trajectories for 12-5 correlation power filter (data windowed).

each end of the interval (as we would expect from our discussion in Section 2).

This example shows that for any of the time-varying methods discussed in this paper, 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 appears 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 from this point on 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 were compared with the time-invariant LPC poles.

The trajectories of the center frequencies shown in Fig. 10 for both methods agree favorably. The main deviations between the time-varying method and regular LPC occurred in the first and second poles at the beginning of the time interval, where the 'lowpass' nature of the time-varying LPC method is most evident. The time-varying method corresponded to 'smoothed' values of the center frequency locations of regular LPC. The radius trajectories of the poles shown in Fig. 11 agree fairly well except for the fifth pole. The center frequency trajectory of the fifth pole matched very well, while the radius trajectory did not. The radius trajectory deviations seem to be a result of the 'lowpass' nature of the time-varying method.

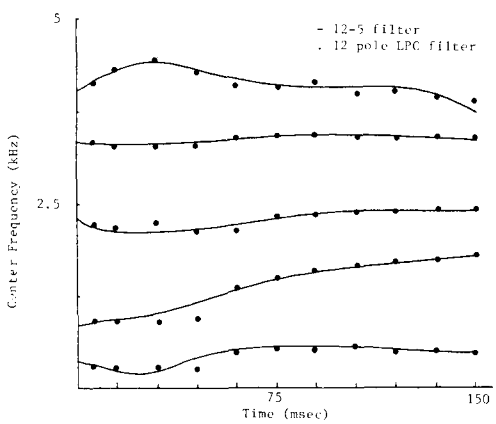


Fig. 10. Center frequency trajectories for 12-5 covariance power filter and 12 pole LPC filters.

Next we compared the log spectra of the all-pole time-invariant and time-varying filters with log spectra of the speech signal. The spectra were compared because LPC can be thought of as attempting to match the spectral envelope of speech with the spectrum of the all-pole filter. This is discussed in detail in [2]. For the time-varying case, the spectrum was defined at a time instant k as the frequency response of the filter with coefficients $a_i(k)$ for $i = 0, \ldots, p$.

The spectra for the regular LPC and time-varying LPC filters for selected times are shown in Fig. 12. The spectra have been adjusted so that the largest value is 0 dB.

We used a log spectral measure to determine quantitatively the difference between the spectra for both LPC methods [8, 9]. Following the derivation given by Turner and Dickinson, 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

$$(d_2)^2 = \int_{-\pi}^{\pi} |\ln(G^2/|A(e^{j\theta})|^2) - \ln(G^2/|A'(e^{j\theta})|^2)| \frac{d\theta}{\pi}. \tag{5.1}$$

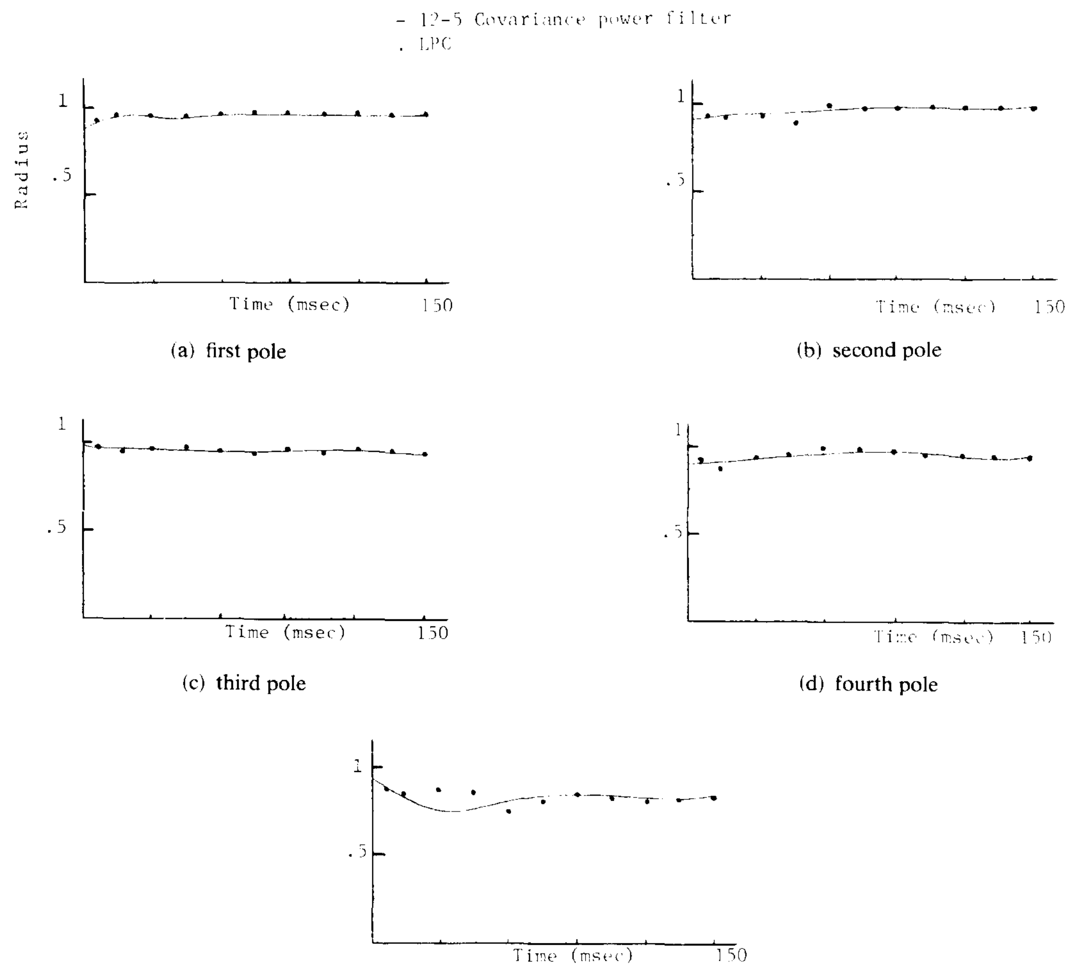
The Taylor series expansion for $\ln A(z)$ (assuming $A(z)$ is stable) is

$$\ln A(z) = -\sum_{k=1}^{\infty} c_k z^{-k} \tag{5.2}$$

with the cepstral coefficients given by

$$c_0 = \ln(G^2),$$

$$c_k = -a_k - \sum_{n=1}^{k-1} c_{k-n} a_n \frac{(k-n)}{n}, \quad k > 0. \tag{5.3}$$



(a) first pole (b) second pole

(c) third pole (d) fourth pole

(e) fifth pole

Fig. 11. Radius trajectories.

By applying Parseval's relationship to (5.3), the log spectral measure is

$$(d_2)^2 = \sum_{k=-\infty}^{\infty} (c_k - c'_k)^2 \tag{5.4}$$

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

$$\mathrm{SPDIFF} = \left(\frac{10}{\ln 10}\right) \cdot 2\left[\sum_{i=1}^{p} (c_k - c'_k)^2\right]^{1/2}. \tag{5.5}$$

Markel and Gray [8] have reported that there is a high correlation between SPDIFF and d_2. Turner and Dickinson [9] 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 have also developed an average SPDIFF for filters with time-varying coefficients. In our study, 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

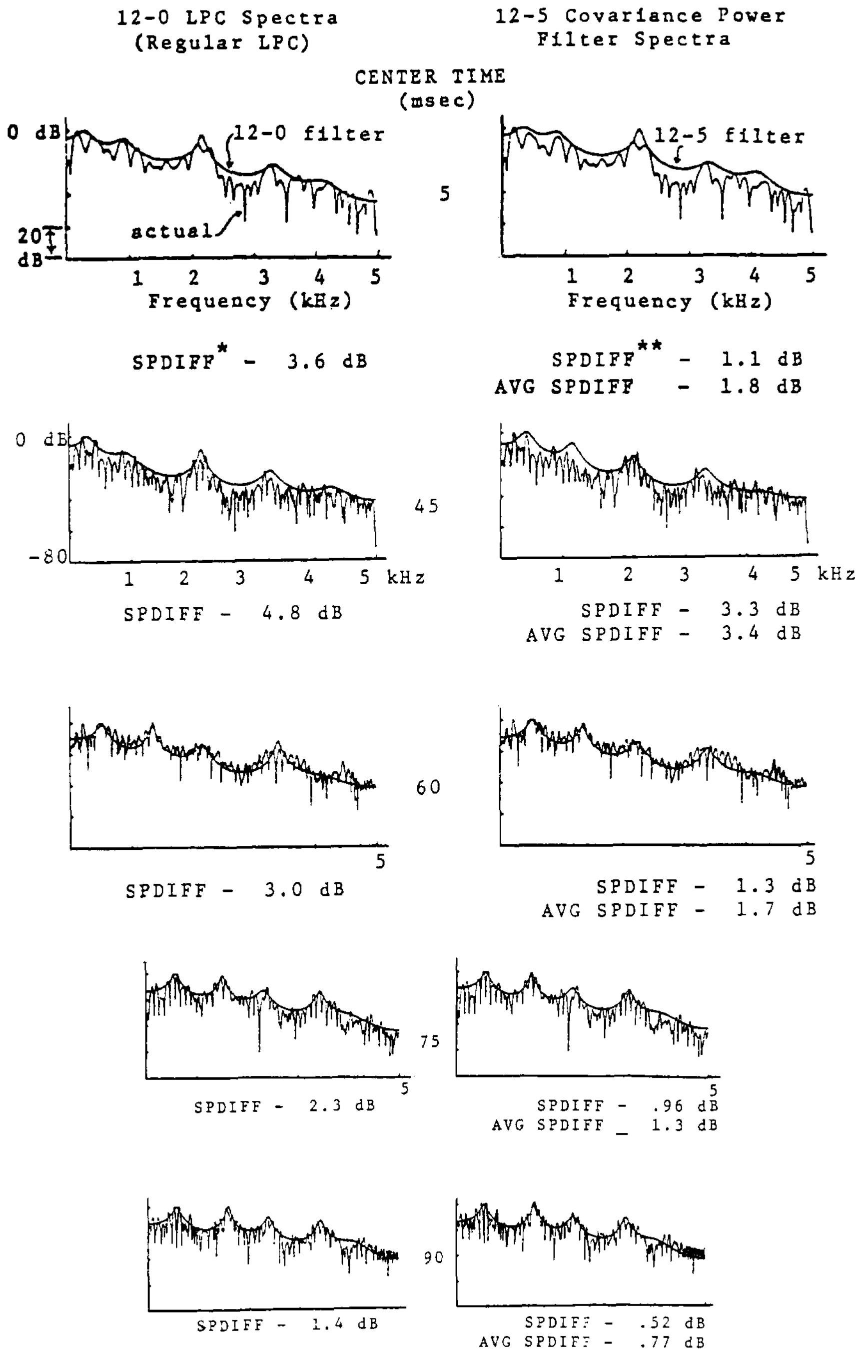


Fig. 12. Comparison of actual and filter spectra for pre-emphasized speech example. * Difference between 12-0 (regular) LPC spectra for successive center times (i.e., between 5 and 15 msec). ** Difference (and average difference) between 12-0 and 12-5 filter spectra for same center time.

assume to be $[1, L]$). For this, the time-average spectral difference is

$$\text{AVG SPDIFF} = \left(\frac{10}{\ln 10}\right) \cdot \left[\frac{1}{L} \sum_{n=1}^{L} 2 \sum_{k=1}^{P} (c_k - c_k'(n))^2\right]^{1/2} \tag{5.6}$$

where the cepstral coefficients are $c_k(n)$ 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 the right-hand column of Fig. 12. The time average of the spectral difference, AVG SPDIFF, 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. 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 in the left-hand column.

There are large spectral differences between the successive regular LPC time-invariant filters for the comparison times of 45 and 60, and 60 and 75 msec. These are the times in which the signal characteristics change significantly. The largest average spectral differences between the time-varying LPC filter and the regular LPC time-invariant filters occur at the times of 30 and 45 msec. The values of the average spectral differences were 2.5 and 3.4, respectively, which would indicate that the differences between the two methods would be perceptible. After 60 msec, the average differences 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 45 msec can be explained in part because of the 'lowpass' action of the time-varying filter. 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. 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. This is discussed further in [4].

As an example of the effect of not estimating the time-varying gain, we attempted to reproduce the original pre-emphasized signal. For this example, 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 Fig. 13.

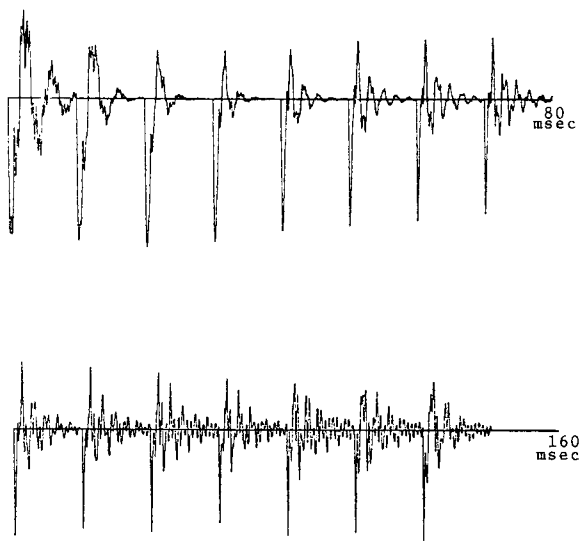


Fig. 13. Reproduction of original signal using 15-4 covariance power filter.

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 Fig. 7 are there.

This 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 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.

6. Discussion

In this paper we have developed a method of time-varying linear prediction for the analysis of nonstationary speech signals. In our approach the coefficients of the speech production system are modeled as linear combinations of a set of known time functions. The coefficients in these linear combinations can be determined by solving a set of linear equations, much as in standard LPC, and we have briefly discussed the structure and efficient solution of these equations.

Perhaps the most important contribution of this paper is the investigation of several methods for the evaluation of the performance of time-varying identification methods in general. By applying these methods to time-varying LPC we have been able to uncover some of its basic properties and to gain some insight into issues that may arise in time-varying modeling. Specifically, we have seen that by modeling parameter time variations in a smooth fashion (as in the power or Fourier methods), the resulting system tends to respond in a low-pass fashion to abrupt signal changes. Therefore, this method is most effective in tracking slowly varying signal characteristics. If we are interested in detecting or tracking abrupt changes, several possibilities present themselves. We have seen that traditional time-invariant LPC performed over shorter analysis intervals can track abrupt changes, at the expense of an increased number of parameters. It is interesting to note that time-invariant LPC over a sequence of intervals is a special case of time-varying LPC in which we use basis functions of the form

$$u_k(n)=\begin{cases}1, & (k-1)N \leqslant n < kN,\\ 0, & \text{otherwise,}\end{cases}$$

so that the coefficients $a_i(n)$ are piecewise constant over intervals of length N. Thus, by choosing discontinuous basis functions we can model abrupt changes. As we observed in our study of time-varying LPC to a speech waveform, abrupt changes in the data can lead to modeling errors if a smooth parameter behavior is assumed. Perhaps a combination of smoothly and abruptly varying basis functions will turn out to perform well.

In applying time-varying LPC to speech data, a number of additional characteristics and limitations of the method were uncovered. We saw that the time-varying 'poles' of the estimated model are not guaranteed to remain within the unit circle. The probability of this occurring is reduced if the data is not windowed, but the possibility remains. It may be possible to develop a time-varying estimation method or to determine sets of basis functions for time-varying LPC that will necessarily lead to stable filters. This remains for the future.

For the speech example, the time-varying filter 'tracked' the parameter 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.

Also, as we have discussed, time-varying LPC does not directly produce a time-varying gain estimate. A procedure for determining such an estimate is clearly needed. Perhaps a method could be developed that would both equalize the signal in conjunction with providing a time-varying gain.

It is our feeling that the method of time-varying LPC has promise. It has the potential advantage of reducing the total number of coefficients needed to model a segment of speech, and it also provides a smoothed trajectory of the formants of the vocal tract. A number of limitations and questions have been raised about the method and about time-varying modeling in general. We have attempted to expose these issues, and it is our hope that our work will provide some useful insights and perspectives for future work on time-varying signal representations.

References

[1] J.I. Makhoul, "Linear prediction: A tutorial review", *Proc. IEEE*, Vol. 32, April 1975, pp. 561–582.

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

[3] L.A. Liporace, "Linear estimation of nonstationary signals", *J. Acoust. Soc. Amer.*, Vol. 58, No. 6, December 1975, pp. 1268–1295.

[4] M.G. Hall, "Time-varying linear predictive coding of speech signals", *S.M. Thesis*, Dept. of Electrical Engineering and Computer Science, Mass. Institute of Tech., Cambridge, MA, August 1977.

[5] 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.

[6] J.J. Cornyn, Jr., "Direct methods for solving systems of linear equations involving Toeplitz or Hankel matrices", *Rept. AD/A-022 931*, prepared for U.S. Office of Naval Research, October 1974; also *S.M. Thesis*, Univ. of Maryland, College Park, MD.

[7] R.A. Wiggins and E.A. Robinson, "Recursive solutions to the multi-channel filtering problem", *J. Geophys. Res.*, Vol. 70, 1965, pp. 1885–1891.

[8] A.H. Gray, Jr and J.D. Markel, "Distance measures for speech processing", *IEEE Trans. Acoust. Speech Signal Process*, Vol. 24, No. 5, October 1976, pp. 380–391.

[9] J. Turner and B. Dickinson, "Linear prediction applied to time-varying all-pole signals", *Proc. 1977 IEEE Int. Conf. Acoust. Speech Signal Process.*, Hartford, CT, May 9–11, 1977, pp. 750–753.

[10] Y. Grenier, "Identification of nonstationary ARMA models", presented at the Colloquium entitled "Fast Algorithms for Linear Dynamical Systems", INRIA 2-7261-0264-6 (Vol. II), Aussois, France, Sept. 21–25, 1981.