A sample selective linear predictive analysis of speech signals

Suresh Kumar Venkata Vishnubhatla

University of Nevada, Las Vegas

Follow this and additional works at: https://digitalscholarship.unlv.edu/rtds

Part of the Higher Education Commons

Repository Citation

https://digitalscholarship.unlv.edu/rtds/224

This Thesis is brought to you for free and open access by Digital Scholarship@UNLV. It has been accepted for inclusion in UNLV Retrospective Theses & Dissertations by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
A sample selective linear predictive analysis of speech signals

Vishnubhatla, Suresh Kumar Venkata, M.S.

University of Nevada, Las Vegas, 1992
A SAMPLE SELECTIVE LINEAR PREDICTIVE
ANALYSIS OF SPEECH SIGNALS

by

Suresh Kumar Vishnubhatla

A thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Electrical and Computer Engineering

Department of Electrical and Computer Engineering
University of Nevada, Las Vegas
August, 1992
The thesis of Suresh Kumar Vishnubhatla for the degree of Master of Science in Electrical and Computer Engineering is approved.

Chairperson, Eugene E. McGaugh, Ph.D

Exmaining Committee Member, Sahjendra Singh, Ph.D

Exmaining Committee Member, Ashok Iyer, Ph.D

Graduate Faculty Representative, Satish Bhatnagar, Ph.D

Graduate Dean, Ronald W. Smith, Ph.D

University of Nevada, Las Vegas

August, 1992
ACKNOWLEDGMENTS

I sincerely thank my Advisor Dr. Eugene E. McGaugh for his direction and the time he spared for me during this work. I also thank Dr. Ashok Iyer, Dr. Sahjendra Singh and Dr. Satish Bhatnagar for consenting to serve on my Graduate Committee.

I thank Dr. William Brogan for clearing most of my inhibitions about Random Processes and Dr. Kyoung Kim for introducing me to this fascinating field of Speech Processing.
Abstract

The Linear Prediction Analysis is one of the popular methods of processing speech. But it has problems in estimating the vocal tract characteristics of voiced sounds uttered by females and children. This is because the conventional linear prediction method assumes that all the sample values in each analysis frame are to be approximated by a linear combination of a definite number of the previous samples whether the previous samples include excitation periods or not. So, it sometimes fails to accurately estimate the vocal tract characteristics of short pitch periods like those uttered by females and children. Also, the Linear Prediction analysis is easily affected by source excitation and hence it sometimes fails to extract system parameters. This is especially true in voiced speech of short pitch periods.

The vocal tract characteristics of signals of short pitch period can be estimated more accurately by the Sample Selective Linear Prediction (SSLP). The SSLP is a two stage linear prediction analysis using only relevant samples in the second stage analysis, whereas the conventional Linear Predictive Analysis (LP) uses all the samples with equal weights. The first stage of a SSLP analysis is the conventional linear predictive analysis and in the second stage, only those samples which are under a specified threshold are used for further analysis.

This work outlines a numerically stable algorithm for performing the SSLP using the Autocorrelation method.
## Contents

Acknowledgements iii

Abstract iv

List of Figures vii

1 Introduction to Linear Predictive Coding of Speech 1
   1.1 Parameter Estimation 4

2 Sample Selective Linear Predictive Analysis of Speech 26
   2.1 LP Formulation Using Generalized Inverse of Matrices 27
   2.2 Sample Selective Linear Prediction 29

3 Analysis of Speech Signals of Short Pitch Period by SSLP 33
   3.1 Introduction 33
   3.2 Expectations from the Modifications 34
   3.3 Model of the System 35
   3.4 Error Measure 42

4 Results and Discussion 44
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Speech Production Model for Linear Predictive Coding</td>
<td>2</td>
</tr>
<tr>
<td>Figure 1.2 (a)</td>
<td>Discrete all-pole model in a frequency domain</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1.2 (b)</td>
<td>Discrete all-pole model in the time domain</td>
<td>4</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Selecting sample points for predicted samples</td>
<td>30</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Computation of autocorrelation coefficient</td>
<td>39</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Speech Sample Segment of a female for the vowel /a/</td>
<td>49</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Reconstructed Speech Signal after SSLP for filter order 4</td>
<td>50</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Reconstructed Speech Signal after LPC for filter order 4</td>
<td>51</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Excitation Signal for Reconstruction</td>
<td>52</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Reconstructed Signal after LPC for filter order 6</td>
<td>53</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Comparison between actual and reconstructed signals</td>
<td>54</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Comparison between Spectral Envelopes of true and SSLP</td>
<td>55</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Comparison between Spectral Envelopes of true and LPC</td>
<td>56</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Comparison between Spectral Envelopes of true and LPC</td>
<td>57</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Comparison between the errors by LPC and SSLP</td>
<td>58</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Comparison between actual and reconstructed signals</td>
<td>59</td>
</tr>
<tr>
<td>Figure 4.12</td>
<td>Comparison between actual and LPC signals</td>
<td>60</td>
</tr>
<tr>
<td>Figure 4.13</td>
<td>Comparison between actual and SSLP signals</td>
<td>61</td>
</tr>
</tbody>
</table>
Chapter 1

INTRODUCTION TO
LINEAR PREDICTIVE
CODING OF SPEECH

The method of linear prediction is widely regarded as one of the most powerful speech analysis techniques. This method has become the predominant technique for estimating the basic speech parameters like pitch, formants, spectra and vocal tract area functions. The importance of this method lies both in its ability to provide extremely accurate estimates of the speech
Figure 1.1: Speech production model for Linear Predictive Coding

parameters, and in its relative speed of computation.

The basic idea behind linear predictive analysis is that a speech sample can be approximated as a linear combination of past speech samples. By minimizing the sum of the squared differences over a finite interval, between the actual speech samples and the linearly predicted ones, a unique set of predictor coefficients can be determined. These predictor coefficients are the weighting coefficients used in the linear combination.

The theory of linear prediction is closely related to the speech synthesis model shown in Figure 1.1.
For the application of the model, it is necessary to determine if the signal is voiced or unvoiced and, if voiced, to compute the pitch period. The main difference between LPC and the other vocoders is in the modeling of the vocal tract filter (and secondarily, the computation of the gain $G$). In LPC, the vocal tract is modeled as an all-pole digital filter, i.e., as a filter that only has poles and no zeros. Incorporating a gain $G$ into that filter, we can express it as

$$H(z) = \frac{G}{1 + a_1 z^{-1} + \ldots + a_p z^{-p}} = \frac{S(z)}{U(z)}$$  \hspace{1cm} (1.1)$$

where $p$ is the order of the model. If $s(n)$ is the speech output of the model and $u(n)$ is the excitation input, then equation 1.1 can be written in the time domain as

$$s(n) = Gu(n) - a_1 s(n - 1) - \ldots - a_p s(n - p)$$  \hspace{1cm} (1.2)$$

i.e., every speech sample is computed as a linear combination of the previous speech samples with a contribution from the excitation. Hence, the name *Linear Predictive Coding*. 
1.1 PARAMETER ESTIMATION

In the all-pole model, we assume that the signal $s(n)$ is given by a linear combination of past values and some inputs $u(n)$. This model is shown in Figure 1.2. The problem now is to determine the predictor coefficients $a_p$ and the gain $G$.

1.1.1 Method of Least Squares

Here we assume that the input $u(n)$ is totally unknown, so the signal $s(n)$ can be predicted only approximately from the linearly weighted summation of the past samples. Let this approximation of $s(n)$ be $\tilde{s}(n)$, where
\[ \hat{s}(n) = - \sum_{k=1}^{p} a(k)s(n - k) \]  

(1.3)

Then the error between the actual value \( s(n) \) and the predicted value \( \hat{s}(n) \) is given by

\[ e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^{p} a(k)s(n - k) \]  

(1.4)

\( e(n) \) is also known as the residual. In the method of least squares the parameters \( a(k) \) are obtained as a result of the minimization of the mean or total squared error with respect to each of the parameters. \( s(n) \) can be either a deterministic signal or a sample from a random process.

**Deterministic Signal**

Let the total squared error be \( E \). Then

\[ E = \sum_{n} e^2(n) = \sum_{n} (s(n) + \sum_{k=1}^{p} a(k)s(n - k))^2 \]  

(1.5)
The range of summation over \( n \) in 1.5 and the definition of \( s(n) \) in that range are of importance. Let us first minimize \( E \) without specifying the range. \( E \) is minimized by setting

\[
\frac{\partial E}{\partial a(i)} = 0, \quad 1 \leq i \leq p \tag{1.6}
\]

From 1.5 and 1.6 we obtain the set of equations:

\[
\sum_{k=1}^{p} a(k) \sum_{n} s(n-k)s(n-i) = -\sum_{n} s(n)s(n-i), \quad 1 \leq i \leq p \tag{1.7}
\]

Equations 1.7 are called the normal equations. For any definition of the signal \( s(n) \) 1.7 forms a set of \( p \) equations in \( p \) unknowns which can be solved for the predictor coefficients \( a(k), \quad 1 \leq k \leq p \) which minimize \( E \) in 1.5.

The minimum total squared error, \( E_p \), is obtained by expanding 1.5
and substituting in 1.7, resulting in

\[ E_p = \sum_n s^2(n) + \sum_{k=1}^p a(k) \sum_n s(n)s(n-k) \]  \hspace{1cm} (1.8)

There are two methods by which we can specify the range of summation over \( n \) and thus estimate the parameters.

\textit{a) Autocorrelation method}

In this method we assume that the error in 1.5 is minimized over the infinite duration \(-\infty < n < \infty\). Equations 1.7 and 1.8 then reduce to

\[ \sum_{k=1}^p a(k) R(i-k) = -R(i), \quad 1 \leq i \leq p \]  \hspace{1cm} (1.9)

\[ E_p = R(0) + \sum_{k=1}^p a(k)R(k) \]  \hspace{1cm} (1.10)

where

\[ R(i) = \sum_{n=-\infty}^{\infty} s(n)s(n+i) \]  \hspace{1cm} (1.11)
is the autocorrelation function of the signal \( s(n) \). \( R(i) \) is an even function of \( i \), i.e.,

\[
R(-i) = R(i) 
\]  
(1.12)

The coefficients \( R(i-k) \) form the autocorrelation matrix and hence the name \textit{autocorrelation method}. The autocorrelation matrix is a Toeplitz matrix; i.e., it is symmetric and all the elements along a given diagonal are equal.

However, in most of the practical cases, the signal \( s(n) \) is known over only a finite interval, or we are interested in the signal over only a finite interval. One method to get around this is to multiply the signal \( s(n) \) by a window function \( w_n \) to obtain another signal \( s'(n) \) that is zero outside some interval \( 0 \leq n \leq N - 1 \). i.e.,

\[
s'(n) = \begin{cases} 
  s(n) & 0 \leq n \leq N - 1 \\
  0 & \text{otherwise}
\end{cases} 
\]  
(1.13)
The autocorrelation function is then given by

\[ R(i) = \sum_{n=0}^{N-1-i} s'(n) s'(n + i) \quad i \geq 0 \quad (1.14) \]

The shape of the window \( w_n \) is of importance and is discussed in the later sections.

b) Covariance method

In contrast to the autocorrelation method, here we assume that the error \( E \) in 1.5 is minimized over a finite interval, say \( 0 \leq n \leq N - 1 \). Equations 1.7 and 1.8 then reduce to

\[ \sum_{k=1}^{p} a(k) \varphi_{ki} = -\varphi_{0i}, \quad 1 \leq i \leq p \quad (1.15) \]

\[ E_p = \varphi_{00} + \sum_{k=1}^{p} a(k) \varphi_{0i} \quad (1.16) \]

where
\[ \varphi_{ik} = \sum_{n=0}^{N-1} s(n - i) s(n - k) \]  

(1.17)

is the covariance of the signal \( s(n) \) in the given interval. The coefficients \( \varphi_{ki} \) in 1.15 form a covariance matrix, and therefore the name covariance method. From 1.17 it can be shown that the covariance matrix \( \varphi_{ik} \) is symmetric, i.e., \( \varphi_{ik} = \varphi_{ki} \). Unlike the autocorrelation method, the covariance matrix is not a Toeplitz matrix (elements along a given diagonal are not equal). However, as the interval over which \( n \) varies tends to infinity the covariance method reduces to the correlation method.

1.1.2 Random Signal

If the signal \( s(n) \) is assumed to be sample of a random process, then the error \( e(n) \) in 1.4 is also sample of a random process. In the least squares method, we minimize the expected value of the square of the error. Thus

\[ E = \mathcal{E}(e^2(n)) = \mathcal{E}(s(n) + \sum_{k=1}^{p} a(k) s(n - k))^2 \]  

(1.18)
Applying 1.6 to 1.18, we obtain the normal equations:

\[ \sum_{k=1}^{p} a(k) \mathcal{E}(s(n - k)s(n - i)) = -\mathcal{E}(s(n)s(n - i)), \quad 1 \leq i \leq p \]

(1.19)

The minimum average error is then given by

\[ E_p = \mathcal{E}(s^2(n)) + \sum_{k=1}^{p} a(k) \mathcal{E}(s(n)s(n - k)) \]

(1.20)

Taking the expectations in 1.19 and 1.20 depends on whether the process \( s(n) \) is stationary or nonstationary.

a) Stationary case

For a stationary process \( s(n) \), we have

\[ \mathcal{E}(s(n - k)s(n - i)) = R(i - k) \]

(1.21)
where $R(i)$ is the autocorrelation of the process. Equations 1.19 and 1.20 now reduce to equations identical to 1.9 and 1.10 respectively. The only difference is that here the autocorrelation is that of a stationary process instead of a deterministic signal. For a stationary (and ergodic) process the autocorrelation can be computed as a time average. Different approximations are available for estimating $R(i)$ from a finite known signal $s(n)$. Using this estimate the stationary case gives the same solution for the coefficients $a(k)$ as the autocorrelation method in the deterministic case.

**b) Nonstationary case**

For a nonstationary process $s(n)$, we have

$$\mathbb{E}(s(n-k)s(n-i)) = R(n-k, n-i)$$ (1.22)

where $R(t, t')$ is the nonstationary autocorrelation between times $t$ and $t'$. $R(n-k, n-i)$ is a function of the time index $n$. If we assume that we want to estimate the parameters $a(k)$ at time $n = 0$, then 1.19 and 1.20 reduce to
\[
\sum_{k=1}^{p} a(k) R(-k, -i) = -R(0, -i) \quad (1.23)
\]

\[
E'_{p} = R(0, 0) + \sum_{k=1}^{p} a(k) R(0, k) \quad (1.24)
\]

In estimating the nonstationary autocorrelation coefficients from the signal \( s(n) \), we should note that nonstationary processes are not ergodic, and, therefore, one cannot substitute the ensemble average by a time average. But, for a certain nonstationary processes known as locally stationary processes, it is reasonable to estimate the autocorrelation function with respect to a point in time as a short-time average. Speech signals belong to this class of nonstationary but locally stationary signals.

So, in a manner analogous to the stationary case, we can estimate \( R(-k, -i) \) by \( \varphi_{ik} \) in 1.17. Using this approximation for the nonstationary autocorrelation leads to a solution for the parameters \( a(k) \) in 1.23 that is identical to that given by 1.15 in the covariance method in the deterministic case. It should also be noted that for a stationary signal \( R(t, t') = R(t - t') \) and therefore the normal equations 1.23 and 1.24.
reduce to 1.9 and 1.10.

1.1.3 Gain Computation

It can be recalled that in the least squares method we assumed that the input is unknown. So, it does not make much sense to determine a value for the gain $G$. But 1.4 can be rewritten as

$$ s(n) = - \sum_{k=1}^{p} a(k)s(n-k) + e(n) \quad (1.25) $$

Comparing this with the equation for the all-pole model

$$ s(n) = - \sum_{k=1}^{p} a(k)s(n-k) + Gu(n) \quad (1.26) $$

we can see that the only input signal $u(n)$ that will result in the signal $s(n)$ as output is that where $G u(n) = e(n)$. That is, the input signal is proportional to the error signal. For any other input $u(n)$, the output from the filter $H(z)$ in Figure 1.2 will be different from $s(n)$. However, if we insist that whatever the input $u(n)$, the energy in the output signal must equal that of the original signal $s(n)$, then we can atleast specify the total energy in the input signal. Since the filter $H(z)$ is fixed, it is
clear from the above that the total energy in the input signal $G u(n)$ must equal the total energy in the error signal, which is given by $E_p$ in 1.9 or 1.16, depending on the method we use.

Two types of input are of interest: the deterministic impulse and stationary white noise.

**Impulse input**

Let the input to the all-pole filter $H(z)$ be an impulse or unit sample at $n = 0$, i.e., $u(n) = \delta_{n0}$, where $\delta_{nm}$ is the Kronecker delta. The output of the filter $H(z)$ is then its impulse response $h(n)$, where

$$h(n) = -\sum_{k=1}^{p} a(k)h(n - k) + G\delta_{n0} \quad (1.27)$$

The autocorrelation $\hat{R}(i)$ of the impulse response $h(n)$ has a relationship with the autocorrelation $R(i)$ of the signal $s(n)$. Multiplying 1.27 by $h(n-i)$ and summing over all $n$ gives us
\[ \hat{R}(i) = -\sum_{k=1}^{p} a(k) \hat{R}(i-k), \quad 1 \leq |i| \leq \infty \] (1.28)

\[ \hat{R}(0) = -\sum_{k=1}^{p} a(k) \hat{R}(k) + G^2 \] (1.29)

Given the condition that the total energy in \( h(n) \) must equal that in \( s(n) \), we must have

\[ \hat{R}(0) = R(0) \] (1.30)

since the zeroth autocorrelation coefficient is equal to the total energy in the signal. From 1.30 and the similarity between 1.29 and 1.28 we can conclude that

\[ \hat{R}(i) = R(i), \quad 0 \leq i \leq p \] (1.31)

This says that the first \( p + 1 \) coefficients of the autocorrelation of the
impulse response of \( H(z) \) are identical to the corresponding autocorrelation coefficients of the signal. Thus, the problem of linear prediction using the autocorrelation method can be stated in a new way as follows. To find a filter of the form \( H(z) \) in 1.1 such that the first \( p + 1 \) values of the autocorrelation of its impulse response are equal to the first \( p + 1 \) values of the signal autocorrelation, and such that 1.28 applies.

From 1.29, 1.31 and 1.10, the gain is equal to

\[
G^2 = E_p = R(0) + \sum_{k=1}^{p} a(k) R(k)
\]  

where \( G^2 \) is the total energy in the input \( G \delta_{n0} \).

White noise input

Let the input \( u(n) \) be a sequence of uncorrelated samples (white noise) with zero mean and unit variance, i.e., \( \mathcal{E}(u(n)) = 0 \), all \( n \) and \( \mathcal{E}(u(n)u(m)) = \delta_{nm} \). Let the output of the filter be \( \hat{s}(n) \). For a fixed filter \( H(z) \) the output \( \hat{s}(n) \) forms a stationary random process:
\[ \hat{s}(n) = -\sum_{k=1}^{p} a(k)\hat{s}(n - k) + Gu(n) \] (1.33)

Multiplying 1.33 by \( \hat{s}(n - i) \), taking the expected values and by noting that \( u(n) \) and \( \hat{s}(n - i) \) are uncorrelated for \( i \neq 0 \), the above result can be shown to be identical to 1.28 and 1.29 where \( \hat{R}(i) = E(\hat{s}(n) \hat{s}(n - i)) \) is the autocorrelation of the output \( \hat{s}(n) \). Therefore, we can conclude that 1.28 and 1.29 completely specify an all-pole random process as well.

For the random case we require that the average energy (or variance) of the output \( \hat{s}(n) \) be equal to the variance of the original signal \( s(n) \), or \( \hat{R}(0) = R(0) \), since the zeroth autocorrelation of a zero-mean random process is the variance. So, by a similar reasoning as given in the previous section we can conclude that 1.31 and 1.32 also apply for the random case.

From the preceding, we can see that the relation linking the autocorrelation coefficients of the output of an all-pole filter are the same
whether the input is a single impulse or white noise. This is because both types of input have identical autocorrelations and identical flat spectra. This dualism between the deterministic and the statistical white noise is useful when modeling the speech process, where both unit impulses as well as the white noise are actually used to synthesize speech.

1.1.4 Computation of Predictor Parameters

In each of the two formulations of linear prediction given in the previous section, the predictor coefficients $a(k), 1 \leq k \leq p$, can be computed by solving a set of $p$ equations with $p$ unknowns. These equations are 1.9 for the autocorrelation (stationary) method and 1.15 for the covariance (nonstationary) method. There are several standard methods for performing the necessary computations, e.g., the Gauss reduction or elimination method. These methods require $p^3/3 + O(p^2)$ operations (multiplications or divisions) and $p^2$ storage locations. It can also be noted from 1.9 and 1.15 that the matrix of coefficients in each case is a covariance matrix. Covariance matrices are symmetric and in general positive semidefinite, although in most of the cases they are
positive definite. So, 1.9 and 1.15 can be solved more efficiently by the square-root or Cholesky decomposition method. This method requires about half the computation \( p^3/6 + O(p^2) \) and about half the storage \( p^2/2 \) of the general methods. This method is considered quite stable numerically.

Further reduction in storage and computation time is possible in solving the autocorrelation normal equations 1.9 because of their special form. Equation 1.9 can be expanded in matrix form as

\[
\begin{array}{cccccccccccc}
R_0 & R_1 & R_2 & \ldots & R_{p-1} & a_1 & R_1 \\
R_1 & R_0 & R_1 & \ldots & R_{p-2} & a_2 & R_2 \\
R_2 & R_1 & R_0 & \ldots & R_{p-3} & a_3 & R_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
 \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
R_{p-1} & R_{p-2} & R_{p-3} & \ldots & R_0 & a_p & R_p \\
\end{array}
\]

(1.34)
It can be seen that the $p \times p$ autocorrelation matrix is symmetric and the elements along any diagonal are identical (i.e., a Toeplitz matrix). Levinson derived a recursive procedure for solving this type of equation. The procedure was later reformulated by Robinson. Levinson's method assumes the column vector on the right hand side of 1.34 to be a general column vector. By making use of the fact that this column vector comprises the same elements found in the autocorrelation matrix, another method derived by Durbin is twice as fast as Levinson's. The Durbin's method requires only $2p$ storage locations and $p^2 + O(p)$ operations. Durbin's recursive procedure can be specified as follows:

\[ E_0 = R(0) \]  \hspace{1cm} (1.35)

\[ k_i = - \left[ R(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j) \right] / E_{i-1} \]  \hspace{1cm} (1.36)

\[ a_i^{(i)} = k_i \]  \hspace{1cm} (1.37)

\[ a_j^i = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)} \quad 1 \leq j \leq i - 1 \]  \hspace{1cm} (1.38)

\[ E_i = (1 - k_i^2) E_{i-1} \]  \hspace{1cm} (1.39)
Equations 1.35-1.39 are solved recursively for $i = 1, 2, \ldots, p$. The final solution is given by

$$a_j = a_j^p, \quad 1 \leq j \leq p \quad (1.40)$$

It can be observed that in obtaining the solution for a predictor of order $p$, we actually compute the solutions for all predictors of order less than $p$. It should be mentioned that the solution of the normal equations 1.9 or 1.15 does not form the major computational load. The computation of the autocorrelation or covariance coefficients require $pN$ operations, which can dominate the computation time if $N >> p$, as is often the case.

The solution to 1.34 is unaffected if all the autocorrelation coefficients are scaled by a constant. In particular, if all $R(i)$ are normalized by dividing by $R(0)$, we get what are known as the normalized autocorrelation coefficients $r(i)$:

$$r(i) = \frac{R(i)}{R(0)} \quad (1.41)$$
which have the property that $|r(z)| < 1$.

The intermediate quantities $k_i, \ 1 \leq i \leq p$, are known as the reflection coefficients.

1.1.5 Filter Stability

Filter stability is important for many applications. A causal all-pole filter is stable if all its poles lie inside the unit circle (in which case it is also a filter with minimum phase). The poles of $H(z)$ are simply the roots of the denominator polynomial $A(z)$, where

$$A(z) = 1 + \sum_{k=1}^{p} a(k) z^{-k} \quad (1.42)$$

and

$$H(z) = \frac{G}{A(z)} \quad (1.43)$$

$A(z)$ is also known as the inverse filter.

If the coefficients $R(i)$ in 1.9 are positive definite (which is assured if $R(i)$ is computed from a nonzero signal using 1.14), the solution of the
autocorrelation equation 1.9 gives predictor parameters which guarantee that all the roots of $A(z)$ lie inside the unit circle, i.e., a stable $H(z)$. The positive definiteness of $R(i)$ can often be lost if one uses a small word length to represent $R(i)$ in a computer. Also, roundoff errors can cause the autocorrelation matrix to become ill-conditioned. Therefore it is often necessary to check for the stability of $H(z)$. Checking if the roots of $A(z)$ are inside the unit circle is a costly procedure. Another method is to check if all the successive errors are positive. The condition $E_i > 0, \ 1 \leq i \leq p$, is a necessary and sufficient condition for the stability of $H(z)$. From 1.39 it is clear that an equivalent condition for the stability of $H(z)$ is that

$$|k_i| < 1, \quad 1 \leq i \leq p$$

(1.44)

Therefore, the recursive procedure 1.35-1.39 also facilitates the check for the stability of the filter $H(z)$.

The predictor parameters resulting from a solution to the covariance matrix equation 1.15 cannot in general be guaranteed to form a stable
filter. The computed filter tends to be more stable as the number of signal samples $N$ is increased, i.e., as the covariance matrix approaches an autocorrelation matrix. A question always arises as to whether to use the autocorrelation method or covariance method in estimating the predictor parameters. The covariance method is quite general and can be used with no restrictions. The only problem is that of the stability of the resulting filter. In the autocorrelation method, on the other hand, the filter is guaranteed to be stable, but problems of parameter accuracy can arise because of the necessity of windowing the time signal. This is usually a problem if the signal is a portion of an impulse response. If the impulse response of an all-pole filter is analyzed by the covariance method, the filter parameters can be computed accurately from only a finite number of samples of the signal. Using the autocorrelation method, one cannot obtain the exact parameter values unless the whole infinite impulse response is used in the analysis. But it has been found in practice that a very good approximation can be obtained by truncating the impulse response at a point where most of the decay of the response has already occurred.
Chapter 2

SAMPLE SELECTIVE LINEAR PREDICTIVE ANALYSIS OF SPEECH

Even though Linear Prediction (LP) analysis is one of the most prevailing methods for speech analysis, its performance is easily affected by source excitation and hence it sometimes fails to extract system parameters. A new method called Sample Selective Linear Prediction (SSLP) was suggested in 1982 [R. Mizoguchi et al]. The basic idea of
SSLP is based on selective use of data samples in the time domain, and is designed to avoid the effect of source characteristics. But it was however applicable only to analysis of stationary parts of voiced sounds because of its excess sensitivity to temporal changes of the vocal tract characteristics.

2.1 LP FORMULATION USING GENERALIZED INVERSE OF MATRICES

The linear prediction model assumes that the i-th sample $S_i$ is predicted by a linear combination of preceding $(i-1)$ samples. Therefore, a set of linear equations of LP model written in the matrix form is as follows [M. Yanagida et al, 1980]:

$$S_{nXp} \cdot a_{pX1} = s_{nX1} \quad (2.1)$$

where

$$a = [a_1 \ a_2 \ \ldots \ a_p]^T \quad (2.2)$$
PLEASE NOTE:

Page(s) not included with original material and unavailable from author or university. Filmed as received.
notice three advantages:

1. It can handle singular cases in the same way as in full rank cases.

2. It clearly shows the point-wise frame shifting.

3. Provides the means for selective use of data samples.

The solution of the above generalized matrix formulation can be obtained by the triangularization method, using an orthogonal matrix. This method is known as Given’s reduction and various algorithms like Gentleman’s algorithm are available to solve the equations recursively.

2.2 SAMPLE SELECTIVE LINEAR PREDICTION (SSLP)

2.2.1 The Concept

SSLP is an analysis method designed so as to avoid the effect of excitation source from analysis. The basic idea of SSLP is that the effect of excitation source can be removed by discarding those sample data of large prediction error out of analysis (See Fig. 2.1). In order to determine which data are to be discarded, the residual signal \( e(t) \) (calcula-
Figure 2.1: An example of the selecting sample points for predicted samples. 
($T_a = 15 \text{ ms}, T_w = 10 \text{ ms}, \theta = 0.5, T_o = 0.5 \text{ ms}.$)

tion of which will be explained in the following sections) is investigated to ascertain whether a particular speech sample $s(t)$ contains such large prediction error that exceeds a certain threshold $\theta$, which is set in advance. Thus, SSLP selects “clean” data in its successive computation to improve the analysis performance.

As it can be noticed, the above outline of the algorithm doesn’t provide a “decision” to distinguish between the prediction error caused by the change of the vocal tract characteristics and that by the glottal excitation. So, there is a need to provide a decision algorithm for SSLP to distinguish between the data affected by slow system transition and
those affected by excitation. So, let us define residual difference as the difference between the residuals at two successive sample points, as it is expected to be sensitive only to the source excitation but not to the slow change of the vocal tract characteristics. So, if the residual difference is used in the decision algorithm instead of the residual itself, the SSLP becomes applicable to transient parts of speech without changing the basic algorithm of SSLP.

2.2.2 Improved SSLP

An improved SSLP method can be the prototype SSLP with a decision algorithm as to which samples are to be retained and which samples to be discarded. We start the analysis by fixing the starting point of the analysis frame, the frame length, prediction order and the discarding range. The preliminary analysis is performed by the normal LP and the excitation point is fixed. Then we fix the threshold for the allowed error. All the samples whose residual error exceeds this threshold value are discarded. The Givens' reduction is then applied to the generalized matrix formulation as explained in the previous sections. The analysis frame is now shifted and the starting points and the threshold are
updated. The above procedure is continued till we reach a point where the error is within our prescribed limits.[Kakusho et al, 1984].

2.2.3 Short-term SSLP

The frame of analysis as we know is not statistically well built-up in the beginning stage of processing, the criterion for sample selection is unstable and is easily affected by particular data samples processed at that stage. Therefore SSLP cannot be applied to short-term analysis. One way to make the working area stable is to perform Givens' reduction twice and evaluate the residual in the second reduction stage.

2.2.4 Summary

The prototype SSLP is a single-stage analysis using the Givens' reduction as the convenient technique of successive processing for solving an overdetermined set of linear equations putting a binary weight on each equation on the way of processing. It has been found that SSLP can be used for the analysis of the non-stationary parts of speech by employing residual difference as reference and a special scheme was also formulated for short-term analysis [Kakusho et al, 1984].
Chapter 3

ANALYSIS OF SPEECH SIGNALS OF SHORT PITCH PERIOD BY SSLP

3.1 Introduction

The prototype SSLP which was described in the previous chapter is a single stage analysis using Givens' reduction of successive processing for solving the set of generalized matrix equations which lead to a set of predictor coefficients. Here we use the modified SSLP to be a kind
of two-stage linear prediction analyses evaluating only such prediction equations that yield relatively small prediction errors (as we specify) while the conventional linear prediction as described in Chapter 1 uses all the prediction equations with equal weights.

3.2 Expectations from the Modifications

After some preprocessing, the excitation waveform in voiced speech can be considered as a train of pseudoperiodic impulses, then a linear prediction analysis free from ill effects caused by glottal excitation can be expected if the SSLP would discard prediction equations which yield large prediction errors from the set of prediction equations. These prediction equations are selected by referring the residual signals obtained through the inverse filter with coefficients calculated by the conventional linear prediction method in the first-stage analysis. We have two choices regarding the method of calculation of the predictor coefficients (Chapter 1): the autocorrelation method and the covariance method. The correlation method will be chosen for numerical stability.

The various stages in the SSLP analysis will be: The conventional
linear predictive analysis will be performed on the speech samples using
the autocorrelation method. The residual signal between the original
samples and the predicted samples will be calculated sample by sample.
The residual signal is then normalized for implementing the decision
algorithm. A threshold \( \theta \) will be specified. \( \theta \) will specify the range
upto which we will retain the predicted samples and all the samples
whose error is greater than \( \theta \) will be discarded along with the transient
samples. The conventional linear prediction is again performed on this
new set of samples to obtain a new set of predictor coefficients and the
speech samples are reconstructed using the new set of predictor coeffi-
cients. This new reconstructed signal will be the SSLP reconstructed
speech signal.

3.3 Model of the System

As explained in Chapter 1, the speech production model can be gener-
ally assumed as an all-pole model represented by the following equation:

\[
s(n) = \sum_{k=1}^{p} a(k) s(n - k) + u(n)
\]  

(3.1)
where \(s(n)\) denotes the \(n\)th sample of a speech wave. \(u(n)\) is the \(n\)th sample of the excitation wave and \(a(k)\) the \(k\)th predictor coefficient. Let the speech signal from the LPC model be \(\tilde{s}(n)\). For a complete representation of the LPC model, the vocal tract filter parameters (i.e., the filter coefficients \(a(k)s\) and the gain \(G\) must be determined. To do that, we set

\[
\tilde{s}(n) = -a(1)s(n-1) - \ldots - a(p)s(n-p) \tag{3.2}
\]

for a predictor of order \(p\), to be the estimate of \(s(n)\) from the previous samples, and we determine the coefficients \(a(k)\), so that error

\[
\sum_{n} (s(n) - \tilde{s}(n))^2 \tag{3.3}
\]

is minimized over all the available samples. Minimization of the total squared error with respect to \(a(k)\) leads to the following set of linear equations:
\[ a(1) r(0) + a(2) r(1) + \ldots + a(p) r(p-1) = -r(1) \]

\[ a(1) r(1) + a(2) r(0) + \ldots + a(p) r(p-2) = -r(2) \]

\[ a(1) r(p-1) + a(2) r(p-2) + \ldots + a(p) r(0) = -r(p) \]

Or in the matrix form,

\[ R \cdot a = -r \]

(3.5)

where

\[ r^T = \begin{bmatrix} r(1) & r(2) & \ldots & r(p) \end{bmatrix} \]

(3.6)

\[ a^T = \begin{bmatrix} a(1) & a(2) & \ldots & a(p) \end{bmatrix} \]

(3.7)
shown in Figure (3.1).

This formulation is the *autocorrelation method* and produces a matrix $R$ that is a Toeplitz matrix. A Toeplitz matrix is one whose diagonals are composed of identical elements. Such a matrix is nonsingular and it can always be inverted. Hence, we can always find a solution

$$a = -R^{-1} r \quad (3.10)$$

Besides this method, as mentioned in the previous sections we can also use the *covariance method*. In the covariance method, the speech signal $s(n)$ is not windowed and instead of the autocorrelations $r(i)$, you compute the covariances $r(i,j)$ for the $(i,j)$ element of the matrix $R$. The covariance $r(i,j)$ is computed from

$$r(i,j) = \sum_n s(n + i) \ s(n + j) \quad (3.11)$$

Now the matrix $R$ is not guaranteed to be invertible. It is possible that the above system of equations does not have a solution. In which case
Figure 3.1: Computation of the autocorrelation coefficient \( r(3) \). (a) signal \( s(n) \); (b) shifted signal \( s(n+3) \); (c) product \( s(n) s(n+3) \) to be summed in order to compute \( r(3) \)
the LPC filter is unstable. For that reason we use the autocorrelation method and not the covariance method in our formulation.

The solution to the set of equation (3.4) and (3.5) can be found by using one of the classical methods of numerical analysis, such as Gauss Elimination. This process gives us the filter coefficients $a(i)$ as

$$ a = -R^{-1} r $$

(3.12)

However since $R$ is a Toeplitz matrix, we have a very efficient way of obtaining the solution to that equation by Durbin's recursive method. In Durbin's method, we start with the autocorrelation coefficients $r(i)$, $i = 0,...,p$ and compute recursively the filter coefficients, $a(i)$, from the following relations:

$$ K(i) = \frac{r(i) + a_1^{i-1}r(i-1) + \ldots + a_{i-1}^{i-1}r(1)}{E(i-1)} \text{ for } i = 1,...,p $$

(3.13)

$$ a_i^{(i)} = K(i) $$

(3.14)
\[ a_j^{(i)} = a_j^{(i-1)} + K(i)a_{i-j}^{(i-1)} \quad 1, \ldots, i - 1 \quad (3.15) \]

\[ E(i) = (1 - K^2(i)) E(i - 1) \quad (3.16) \]

The coefficients \( a_j^{(i)}, j = 1, \ldots, i \) are the filter coefficients of an \( i \)th order model. Hence the coefficients of the desired \( p \)th order model are:

\[ a(j) = a_j^p, \quad j = 1, \ldots, p \quad (3.17) \]

Durbins's solution gives the parameters \( K(i), i = 1, \ldots, p \) and \( E(p) \) as a side product. \( E(p) \) is the square of the gain \( G \) needed in the synthesis model:

\[ G^2 = E(p) \quad (3.18) \]

This quantity can be encoded as one of the necessary parameters for synthesis. However, since
\[ E(p) = (1 - K^2(1))(1 - K^2(2)) \cdots (1 - k^2(p))r(0) \]  

(3.19)

Instead of \( E(p) \) we can encode and transmit \( r(0) \), which is the energy of the speech frame analyzed. Then \( G \) is recovered by (3.19) during synthesis. This is better because the synthesis model is less sensitive to the quantization noise of \( r(0) \) than that of \( G \).

### 3.4 ERROR MEASURE

The normalized mean squared error will be used as the error measure to compare the Conventional Linear Prediction with the proposed Sample Selective Linear Prediction. In both the cases the autocorrelation method will be used. The error measure is defined as

\[
V_n = \frac{\sum_{m=0}^{N+p-1} e_n^2(m)}{\sum_{m=0}^{N-1} \delta_n^2(m)}
\]

(3.20)

where \( e_n(m) \) is the output of the prediction error filter correspond-
ing to the speech segment $s_n(m)$ located at time index $n$. 
Chapter 4

RESULTS AND DISCUSSION

In this chapter results are presented comparing the results obtained by using Sample Selective Linear Prediction with those from the conventional Linear Predictive Analysis.

4.1 Speech Reconstruction

The accuracy of the proposed Sample Selective Linear Prediction is examined on three different speech signals; two female samples and a male sample. The Linear Predictive Analysis and the Sample Selective
Linear Predictive Analysis are performed on the same signal. The sampling rate was 8 KHz. and the analysis frame was 160 samples. The threshold for SSLP was fixed at 0.7. The original speech signal is shown in Figure (4.1). This is a female speech sample of the vowel /a/ in steady state. The sampling of the original speech signal was performed using the Chimera board A/D converter.

The Linear Predictive Analysis was performed on this signal along with the Sample Selective Linear Predictive Analysis. The analysis was varied for filter order 2 to 17 and the optimum filter order for the SSLP was found to be 4 following the error criterion discussed in the previous chapter. The reconstructed speech signal from the SSLP is shown in Figure (4.2). The reconstruction was performed using Equation (1.2).

The reconstructed signal from Linear Predictive Analysis for the same filter order 4 is shown in Figure (4.3). It can be seen that the excitation indeed has an effect on the output. The excitation signal is shown in Figure (4.4). The impulses occur every 16 samples which is the pitch period of the original speech signal.
The reconstructed speech signal from Linear Predictive Analysis for filter order 6 is shown in Figure (4.5). This is the minimum filter order where the LPC performs at least as well as SSLP. The reason for the higher filter order, which implies higher amount of computation and lesser data compression in terms of transmission can be explained by the fact that the SSLP eliminates the effect of the excitation source characteristics on the reconstructed signal by the weighting performed. Whereas, the LPC analysis allots equal weights to all the samples.

A comparison between the original speech signal and the SSLP reconstructed speech signal is shown in Figure (4.6). It can be seen that after the initial samples the reconstructed signal is almost an exact representation of the original speech signal. The nonconformity during the initial part is due to the lack of previous samples for accurate prediction. After the initial P (filter order) samples the reconstruction is accurate.
4.2 Spectral Characteristics

To verify whether the natural frequencies of the reconstructed signal match with those of the original speech signal, Spectral Analysis is performed on both the true and the reconstructed (LPC and SSLP) speech signals. Figure (4.7) shows comparison between the Spectral Envelopes of the original speech signal and the SSLP reconstructed speech signal. It can be seen that there is a very accurate match between the natural frequencies (formants) of both the signals. Figure (4.8) shows a comparison between the spectral characteristics of true speech signal and the LPC reconstructed signal for the same filter order 4. LPC at this stage does not perform as well as SSLP. Figure (4.9) shows LPC at filter order 6 and we can see that it is fairly accurate but not as close a match as SSLP.

4.3 Error

The error measure described in the previous section was used and the filter orders were varied from 2 to 17. The error in the reconstructed signal for both LPC and SSLP is compared in Figure (4.10). It can be
seen that SSLP does produce less error and lower optimum filter orders compared to LPC.

Figures (4.11), (4.12) and (4.13) show the comparison between SSLP and LPC reconstructed signals with respect to the original speech signals for another set of speech samples.
Figure 4.1: Speech Sample segment of a female for the vowel /a/
Figure 4.2: Reconstructed Speech Signal after Sample Selective Linear Prediction Analysis for filter order 4
Figure 4.3: Reconstructed Speech Signal after Linear Predictive Analysis for filter order 4
Figure 4.4: Excitation Signal for Reconstruction
Figure 4.5: Reconstructed Speech Signal after Linear Predictive Analysis
for filter order 6
Figure 4.6: Comparison between actual speech signal and SSLP reconstructed speech signal
Figure 4.7: Comparison between the Spectral Envelopes of the original speech signal and the SSLP reconstructed signal for filter order 4
Figure 4.8: Comparison between the Spectral Envelopes of the original speech signal and the LPC reconstructed signal for filter order 4.
Figure 4.9: Comparison between the Spectral Envelopes of the original speech signal and the LPC reconstructed signal for filter order 6.
Figure 4.10: Comparison between the errors produced by LPC and SSLP
Figure 4.11: Comparison between SSLP reconstructed signal and the original speech signal for the female sound /e/
Figure 4.12: Comparison between LPC reconstructed signal and the original speech signal for the female sound /e/
Figure 4.13: Comparison between SSLP reconstructed signal and the original speech signal for the male sound /a/
Chapter 5

CONCLUSION

In this work, we formulated the Sample Selective Linear Predictive Analysis (SSLP) using the Autocorrelation method. This formulation is numerically stable compared to the other known approaches like the Covariance method. The numerical stability of the system comes from the windowing we perform at the start of our analysis. This makes the speech signals which are basically nonstationary, quasi-stationary. Hence, when the matrix equations are formed we are guaranteed to have a full rank for the $R$ matrix, which in turn guarantees a solution to the equations, which are nothing but the filter parameters. As the filter orders increase the reconstructed speech signal does not really
look like a reproduction of the original speech. This is because of the over specification of the filter order. Even then, a solution to the matrix equations is guaranteed by this formulation.

As it can be observed from the results presented in the previous section, the SSLP indeed extracts the speech characteristics better than the conventional Linear Prediction. This method can be extremely useful when estimating characteristics of female and child speech. The pitch periods of female and child speech are short. In the conventional Linear Prediction the excitation characteristics greatly affect the reconstruction process. In this case, since the pitch periods are short, the excitation pulses come in at a greater frequency than other speech signals. This leads to the greater amount of error in the reconstructed speech. In Sample Selective Linear Prediction, we try to minimize the error due to excitation characteristics by weighting the samples, and hence the better performance.
Bibliography


19. Strobach, Peter, "Linear Prediction Theory: A Mathematical