FUNDAMENTALS OF LINEAR PREDICTION

Shivali Srivastava
Department for Electrical and Computer Engineering
Mississippi State University, Mississippi State, MS 39762
srivasta@isip.msstate.edu

ABSTRACT

The linear prediction method provides a robust, reliable and accurate method for estimating the parameters that characterize the linear time-varying system. It is very important tool in digital signal processing[11] because it deals with applications in variety of areas such as speech signal processing, image processing etc. It has become the predominant technique for estimating the basic speech parameters such as pitch, formants, and spectra. The importance of linear prediction lies in it's ability to provide extremely accurate estimates of speech parameters as well as it's relative speed of computation.

The proposed talk covers the linear prediction model which deals with the problem of predicting the values of a stationary random process either forward in time or, backward in time known as forward and backward prediction. I will discuss basic principles of linear prediction analysis and the fundamental problem related with the analysis as applied to speech processing. As linear prediction refers to a variety of formulations of the problem of modeling the speech waveforms, I will discuss few basic formulations of linear prediction analysis and examine the similarities and difference among them. I will then highlight the basic problems related with linear prediction as applied to speech which will lead to PLP analysis, an extension over linear prediction.

1. INTRODUCTION

One of the results of science of estimation theory has been the development of the linear prediction[16] algorithms. This allows us to compute the coefficients of a time-varying filter which simulated the spectrum of a given sound at each point in time. This filter has found uses in many fields, not the least of which is speech analysis and synthesis as well as music. The use of the linear predictor in speech applications allows us to modify speech sounds in many ways, such as changing the pitch without altering the timing, changing timing without changing pitch, or blending the sounds of musical instruments and voices.

LP provides parametric modeling techniques which is used to model the spectrum as an autoregressive process. These parametric models are basically used in compression systems, system identification problems in modern control systems, time series analysis for economic applications, spectral estimation in signal processing, maximum entropy techniques etc. The basis is the source-filter model where the filter is constrained to be an all-pole linear filter.The ideas of linear prediction have been in use under the names of system estimation and system identification. The term system identification is particularly descriptive of LPC methods in that once the predictor coefficients have been obtained, the system can be uniquely identified to the extent that it can be modeled as an all-pole linear system. The appeal of linear prediction as applied to speech, however, is not only it's predictive function but also the fact that it gives us a very good model of the vocal tract which is useful for both theoretical and practical purposes.

These objectives formed the need for the development of linear prediction techniques.

2. BASIC PRINCIPLES

The block diagram of basic model for speech production, appropriate for the linear predictive analysis[1] can be seen in Figure1. In this model, the composite spectrum effects of radiation, vocal tract and glottal excitation are represented by a time-varying digital filter[18] whose steady state system function is of the form

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k \cdot z^{-k}} \quad (1)$$

The system can be excited by an impulse train for voiced speech or a random sequence of unvoiced speech. The pitch period and voiced/unvoiced parameters can be estimated using linear predictive analysis. The speech samples s(n) can be given by using simple difference equation

$$s(n) = \sum_{k=1}^{p} a_k \cdot s(n-k) + G \cdot u(n) \quad (2)$$

The linear predictor with predictor coefficients α_k, and order p is defined as a system whose output is

$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha_k \cdot s(n-k) \quad (3)$$

The system function of this linear predictor is

$$P(z) = \sum_{k=1}^{p} \alpha_k \cdot z^{-k} \quad (4)$$

The prediction error is defined as

$$e(n) = s(n) - \tilde{s}(n) \quad (5)$$

$$e(n) = s(n) - \sum_{k=1}^{p} \alpha_k \cdot s(n-k) \quad (6)$$

Thus the prediction error filter is the output of the system whose transfer function is

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k \cdot z^{-k} \quad (7)$$

Comparison of equations(2) and(5) suggests that if $\alpha_k = a_k$, then $e(n) = G \cdot u(n)$ and in such condition, prediction error filter A(z) will be an inverse filter for the system H(z).

$$H(z) = \frac{G}{A(z)} \quad (8)$$

The basic problem of linear prediction here is to determine a set of predictor coefficients $\{\alpha_k\}$ directly from speech signal in such a manner as to obtain a good estimate of spectral properties of speech signal through the use of equation(8).

3. LP APPROACH

Given a signal, s(n), we seek to model the signal as a linear combination of its previous samples as given in equation(2). The short term average prediction error is defined as

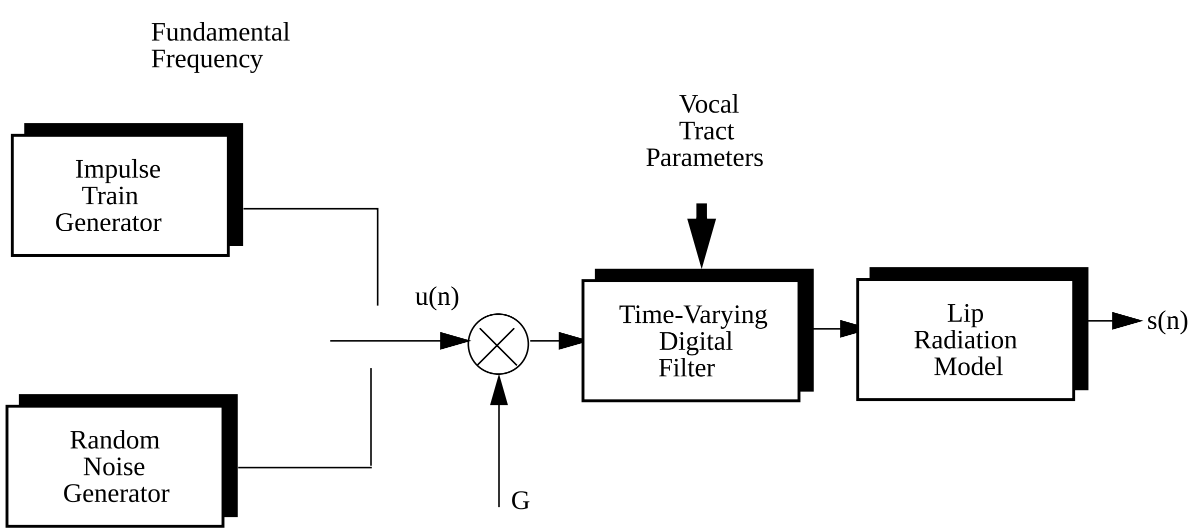

Figure 1. Block diagram of simplified model for speech production

$$E = \sum_n e^2(n) \quad (9)$$

$$E = \sum_n \langle s(n) - \tilde{s}(n) \rangle^2 \quad (10)$$

$$E = \sum_n \left[s(n) - \sum_{k=1}^{p} \alpha_k \cdot s(n-k) \right]^2 \quad (11)$$

We minimize the mean square prediction error[17]. in order to find the values of α_k. This can be done by setting

$$\frac{\partial E}{\partial \alpha_i} = 0 \qquad 1 \le i \le p \quad (12)$$

Differentiation gives

$$\sum_n s(n)s(n-i) = a \quad (13)$$

$$a = \sum_n \left[\sum_{k=1}^{p} \alpha_k s(n-k) \right] \cdot s(n-i) \quad (14)$$

If we define

$$\phi_n(i, k) = \sum_n s(n-i)s(n-k) \quad (15)$$

Then equation(13) and(14) can be more compactly written as

$$\sum_{k=1}^{p} \alpha_k \phi_n(i, k) = \phi_n(i, 0) \quad (16)$$

This is known as linear prediction (Yule-Walker) equation. Solution[19] of this equation gives the values of the predictor coefficients. Mean square prediction error can be written as

$$E = \phi_n(0, 0) - \sum_{k=1}^{p} \alpha_k \phi_n(0, k) \quad (17)$$

Basically two methods are used to solve the linear prediction equation. Here we will consider both of them one by one.

4. THE AUTOCORRELATION METHOD

Constraining the evaluation interval to the range [0, N-1], and assuming the values outside this range to be identically zero, we can take the limits to be [0, N+p-1]. Thus mean square prediction error can be properly expressed as

$$E = \sum_{n=0}^{N+p-1} e^2(n) \quad (18)$$

Because of the finite length constraint, it is important in the autocorrelation method to apply a window. Normally, a Hamming window is used. Application of the window eliminates the problems caused by rapid changes in the signal at the edges of the window, which ensures a smooth transition from frame to frame of the estimated parameters in the overlapping analysis.

Under this condition, we can write equation (15) simply as

$$\phi_n(i, k) = \sum_{n=0}^{N+p-1} s(n-i)s(n-k) \quad (19)$$

which, alternatively, can be expressed as

$$\phi_n(i, k) = \sum_{n=0}^{l} s(n-i)s(n+i-k) \quad (20)$$

$$where \qquad l = N-1-(i-k) \quad (21)$$

Due to finite length constraint we can modify equation (20) as

$$\phi_n(i, k) = R_n(i-k) \quad (22)$$

$R_n(i-k)$ is known as the autocorrelation function evaluated for (i-k). $R_n(i-k)$ is an even function. Under these conditions, linear prediction equation and mean square prediction error can be written, respectively, as

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i) \quad (23)$$

$$E = R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k) \quad (24)$$

In matrix form, the linear prediction equation can be expressed as:

$$\bar{\alpha} = \underline{R}^{-1} \cdot \bar{r} \tag{25}$$

where,

$$\bar{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_p \end{bmatrix} \qquad \bar{r} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \dots \\ R_n(p) \end{bmatrix} \tag{26}$$

$$\underline{R} = \begin{bmatrix} R_n(0) & R_n(1) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & \dots & R_n(p-2) \\ \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & \dots & R_n(0) \end{bmatrix} \tag{27}$$

$\underline{R}$ is symmetric, and all of the elements along the diagonal are equal, which means (1) an inverse exists; (2) the roots are in the left half plane. $\underline{R}$ is known as Toeplitz matrix. This simplification causes the predictor coefficients to be computed efficiently using the Levinson-Durbin recursion[2]:

$$E^{(0)} = R(0) \tag{28}$$

For $1 \le i \le p$ and $1 \le j \le i-1$

$$c_i = \left[R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right] \tag{29}$$

$$k_i = \frac{c_i}{E^{(i-1)}} \tag{30}$$

$$\alpha_i^i = k_i \tag{31}$$

$$\alpha_j^i = \alpha_j^{(i-1)} - k_i \cdot \alpha_{j-1}^{(i-1)} \tag{32}$$

$$E^{(i)} = (1 - k_i^2) \cdot E^{(i-1)} \tag{33}$$

Equations(29) to(33) are solved recursively for i = 1, 2,..., p. and the finally the predictor coefficients are given as

$$\alpha_j = \alpha_j^{(p)} \qquad 1 \le j \le p \tag{34}$$

The complexity of computation in this method of solving predictor coefficients is proportional to p^2, and allow the entire LP computation to be performed with a complexity of approximately $Np + 3N + p^2$.

The signal model, which actually is the inverse of $A(z)$, is given by:

$$H(z) = \frac{G}{A(z)} \tag{35}$$

Where G is the model gain, and is given as:

$$G = \sqrt{E^{(p)}} \tag{36}$$

Which can also be expressed as:

$$G = R_n(0) \prod_{i=1}^{p} (1 - k_i^2) \tag{37}$$

The gain term allows the spectrum of the LP model to be matched to the spectrum of the original speech signal.

The important observations[4] obtained by the use of this form of LP solution are as follows:

(1) The intermediate variables k_i, called the reflection coefficients are bounded:

$$-1 \le k_i \le 1 \tag{38}$$

This result is extremely useful for storage and compression applications involving LP models, such as for the speech recognition system that store large numbers of recognition models[5].

(2) In the process of solving the predictor coefficients of order p using this iterative method, the solutions for the predictor coefficients as well as predictor error of all orders less than p have been obtained. This is convenient for signal processing applications that require estimation of the model order as part of the task.

(3) Model fit becomes better with increase in the order of linear predictor. This fact can easily be verified using equation(33). $E_0 > E_1 > \dots > E_p$. Higher order linear predictor easily represents the finer details in the spectrum, whereas lower order represents only trend of the spectrum.

(4) The reflection coefficients are orthogonal in the sense that the best order "p" model is also the first "p" coefficients in the order "p+1" LP model.

5. THE COVARIANCE METHOD

Fixation of the interval over which mean-squared error is calculated gives a different way to solve the linear predictor equations. We define:

$$E = \sum_{n=0}^{N-1} e^2(n) \tag{39}$$

Thus $\phi_n(i, k)$, for $1 \le i \le p$ and $0 \le k \le p$, can be written as:

$$\phi_n(i, k) = \sum_{n=0}^{N-1} s(n-i)s(n-k) \tag{40}$$

Changing the index of summation simplifies above equation to:

$$\phi_n(i, k) = \sum_{n=0}^{N-i-1} s(n)s(n+i-k) \tag{41}$$

$$\phi_n(i, k) = \sum_{n=0}^{N-k-1} s(n)s(n+k-i) \tag{42}$$

$$\phi_n(i, k) = \phi_n(k, i) \tag{43}$$

Using above equations, the linear predictor equations:

$$\sum_{k=1}^{p} \alpha_k \phi_n(i, k) = \phi_n(i, 0) \tag{44}$$

can be written as:

$$\bar{\phi} = \underline{\Phi} \cdot \bar{\alpha} \tag{45}$$

where:

$$\bar{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_p \end{bmatrix} \qquad \phi = \begin{bmatrix} \phi_n(1, 0) \\ \phi_n(2, 0) \\ \dots \\ \phi_n(p, 0) \end{bmatrix} \tag{46}$$

$$\underline{\Phi} = \begin{bmatrix} \phi_n(1, 1) & \phi_n(1, 2) & \dots & \phi_n(1, p) \\ \phi_n(2, 1) & \phi_n(2, 2) & \dots & \phi_n(2, p) \\ \dots & \dots & \dots & \dots \\ \phi_n(p, 1) & \phi_n(p, 2) & \dots & \phi_n(p, p) \end{bmatrix} \tag{47}$$

The $p \times p$ matrix is a positive definite symmetric matrix and have properties of covariance matrix[7]. For the solution we use Cholesky decomposition method[8], where the matrix $\underline{\Phi}$ is decomposed in the form:

$$\underline{\Phi} = \underline{V} \cdot \underline{D} \cdot \underline{V}^t \tag{48}$$

where V is the lower-triangular symmetric matrix with main diagonal elements equal to 1, D is a diagonal matrix and V^t is the transpose of the matrix V. Equation[48] can be, for $1 \le j \le i$, written as:

$$\phi_n(i, j) = \sum_{k=1}^{j} V_{ik} \cdot d_k \cdot V_{jk} \tag{49}$$

which can be further simplified to:

$$V_{ij} d_j = \phi_n(i, j) - \sum_{k=1}^{j-1} V_{ik} \cdot d_k \cdot V_{jk} \tag{50}$$

Diagonal elements can be written as:

$$\phi_n(i, i) = \sum_{k=1}^{i} V_{ik} \cdot d_k \cdot V_{ik} \tag{51}$$

or, for $i \ge 2$

$$d_i = \phi_n(i, i) - \sum_{k=1}^{i} V_{ik}^2 \cdot d_k \tag{52}$$

The initial condition is:

$$d_1 = 1 \tag{53}$$

Thus we solve for V and D. The linear prediction equation can be written, using notation of equation[48], as:

$$\underline{V} \cdot \underline{D} \cdot \underline{V}^t \cdot \bar{\alpha} = \bar{\phi} \tag{54}$$

Considering

$$\bar{Y} = \underline{D} \cdot \underline{V}^t \cdot \bar{\alpha} \tag{55}$$

linear prediction equation can be written as:

$$\underline{V} \cdot \bar{Y} = \bar{\phi} \tag{56}$$

or

$$\underline{V}^t \cdot \bar{\alpha} = \underline{D}^{-1} \cdot \bar{Y} \tag{57}$$

Y is a column vector which can be solved, for $2 \le i \le p$, using a simple recursion as:

$$Y_i = \phi_i - \sum_{j=2}^{i-1} V_{ij} Y_j \tag{58}$$

with the initial condition:

$$Y_1 = \phi_1 \tag{59}$$

Using vector Y, the linear predictor coefficients can be solved, for $1 \le i \le p-1$, recursively using:

$$\alpha_i = \frac{Y_i}{d_i} - \sum_{j=1}^{p} V_{ji} \alpha_j \tag{60}$$

with initial condition:

$$\alpha_p = \frac{Y_p}{d_p} \tag{61}$$

The mean square prediction error given as:

$$E = \phi_n(0,0) - \sum_{k=1}^{p} \alpha_k \phi_n(0,k) \tag{62}$$

can be simplified as follows:

$$E = \phi_n(0,0) - \bar{\alpha}^t \cdot \phi \tag{63}$$

or

$$E = \phi_n(0,0) - Y^t D^{-1} V^{-1} \cdot \phi \tag{64}$$

which can be further simplified, using equation(59), as:

$$E = \phi_n(0,0) - Y^t D^{-1} Y \tag{65}$$

$$E = \phi_n(0,0) - \sum_{k=1}^{p} \frac{Y_k^2}{d_k} \tag{66}$$

Thus mean square prediction error can be determined directly using the column vector Y and the matrix D. Similar to the Durbin's recursive method, here also the predictor error of all orders less than p is recursively obtained, thereby giving the idea as to how the mean-squared prediction error varies with the order of linear predictor.

6. SPECTRAL ANALYSIS

Addition of power an fundamental frequency information to the LP coefficients allows to reconstruct an audio version of speech signal. LP is used to represent the spectral characteristics of speech[14]. Linear prediction method is preferred over the other methods of spectral analysis, such as, band-pass filter and analysis-by-synthesis for the following reasons:

(1)It provides non-iterative parameter determination for the spectral model

(2)In order to accurately represent the trend characteristic, very small number of parameters are required

(3)A gain constant is easily obtained to match spectral energies of the model and the data, using the autocorrelation method

(4)The model spectrum represents the smoothed version of the data spectrum.

For the spectral analysis of speech, there are two reasons for using linear prediction:

(1)The spectral resonances of voiced speech are weighted most heavily in the error criterion and thus represented most accurately.

(2)The all-pole model can be accurately fit to the log spectrum of a voiced sound with a sufficiently small number of resonances so that the problem of formant extraction reduces to simple peak picking.

The model used for representing the input data spectrum $|X \cdot [\exp(j\theta)]|^2$ is given by:

$$\frac{\sigma^2}{\left|A(e^{j\theta})\right|^2} = \left|\frac{\sigma}{A(z)}\right|^2_{z = e^{j\theta}} \qquad (67)$$

The use of autocorrelation method for spectral estimation is global in that several pitch periods must be contained within the analysis window for meaningful results with voiced speech. As a practical matter, it is generally desirable to use the minimum number of parameters necessary to accurately model the significant features of the signal. In spectral modeling of speech, these features are the vocal tract resonances and regions between these resonances. If “L” is the length of the vocal tract and “c” is the speed of sound, then the memory of the model A(z) must be equal to twice *2L/c*, which is the time required for sound wavs to travel from the glottis to lips. For examples, the representive values 34 cm/sec for c and 17 cm for L result in filter order 10, when sampling rate is 10 kHz and 7, when sampling rate is 6.5 kHz.

7. FORMANT ESTIMATION

Automatic formant analysis[20] is a major problem due to the fact that the vocal tract impulse response is not a directly observable quantity. Parameters of an all-pole are desired where the signal to be processed is the model convolved with a quasi-periodic glottal driving function. For accurate estimation, it is therefore necessary to perform a deconvolution to separate the impulse response and the driving function. Cepstral analysis[10] and linear prediction are the two widely used techniques. LP offers the advantages of minimal complexity, minimum computation time, and maximal accuracy in formant estimation.

A general procedure for formant trajectory estimation based upon linear prediction analysis is shown in Figure2 . Each frame of speech to be analyzed is denoted by the N-length sequence {s(n)}. The speech is first preprocessed by pre-emphasis and possibly windowing. The

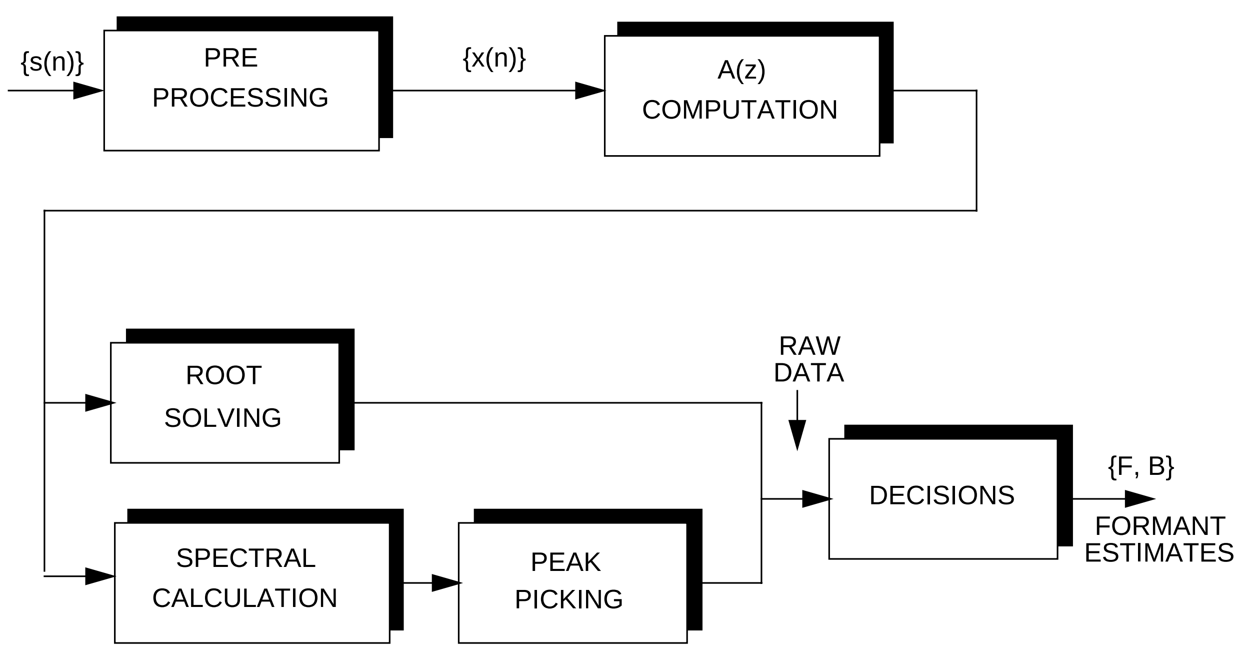

Figure 2. Block diagram of a general procedure for formant energy estimation using linear prediction

preprocessed speech is used to design the inverse filter A(z). Initial estimates for formant frequencies and bandwidths are defined by solving for the roots of polynomial A(z), and searching for the peaks in the spectrum. Solving for the roots guarantees that all possible formant frequency and bandwidth candidates will be extracted. Parabolic interpretation about the peaks results in initial estimates for the formant frequencies and bandwidths $\hat{F}_i$ and $\hat{B}_i$ respectively. These are known as raw data.

The bandwidth $\hat{B}$ and frequency $\hat{F}$ for any complex root z are obtained from the s-plane to z-plane transformation z = exp(sT) where $s = -\pi\hat{B} \pm j \cdot 2\pi\hat{F}$. If z = Re(z) + Im(z) defines the real and imaginary terms of a complex root, then the raw data can be given as:

$$\hat{B} = -(f_s/\pi) \cdot \ln|z| \qquad Hz \tag{68}$$

and

$$\hat{F} = \frac{f_s}{2\pi} \cdot \operatorname{atan}\left[\frac{Im(z)}{Re(z)}\right] \qquad Hz \tag{69}$$

The technique of formant estimation using linear prediction is quite similar to formant estimation using cepstral smoothing. The basic idea is that the frames of windowed speech data are transformed into smoothed spectral representations which have the fundamental frequency associated with the fine-grained structure removed. From the smoothed spectra all peaks are picked as potential candidates for formant indicators.

8. ALL-POLE LP MODEL: ISSUES

LP provides an all-pole filter to model the vocal tract. This technique is quite accurate for vowels and vowel-like sounds. However, the tract introduces both poles and zeros into the speech spectrum. Zeros are manifested by antiresonances in the spectrum of speech sounds. They arise when nasal couples with the oral tract in producing nasalized sounds such as **/m/, /n/, or /ng/**. They can also arise due to the speaker’s environment, or if the excitation source is not at the glottis but in the interior of the vocal tract. The general assumption that zeros occurring in the short term speech spectrum can be approximated by the all-pole filter, is not satisfactory. Although a zero can be approximated arbitrarily closely by a large number of poles.

9. PERCEPTUAL LINEAR PREDICTION

In PLP analysis[9], the all-pole modeling is applied to an auditory spectrum derived by (a) convolving $P(w)$ with a critical band masking pattern, followed by (b) resampling the critical band spectrum at approximately l Bark intervals, (c) pre-emphasis by a simulated fixed equal loudness curve, and finally (d) compression of the resampled and pre-emphasized spectrum through the cubic root non-linearity, simulating the intensity-loudness power law. The low order all-pole model of such an auditory spectrum has been found to be consistent with several phenomena observed in speech perception[3].

The block diagram of PLP Analysis[6] is shown in Figure 3.

After windowing, the real and imaginary components of the short-term speech spectrum are squared and added to get the power spectrum,

$$P(w) = Re[S(w)]^2 + Im[S(w)]^2 \tag{70}$$

9.1. CRITICAL BAND SPECTRAL RESOLUTION

The power spectrum is warped onto a bark scale using approximation:

$$\Omega(w) = 6\ln[(f_1(w)) + [f_2(w)]^{0.5}] \tag{71}$$

$$f_1(w) = w/(1200\pi) \tag{72}$$

$$f_2(w) = [w/(1200\pi)^2 + 1]^{0.5} \tag{73}$$

The bark scale spectra is convolved with the power spectra of the critical band filter. This simulates the frequency resolution of the ear which is approximately constant on the Bark scale.

$$\phi(\Omega) = \begin{cases} 0 & ,\Omega < -1.3 \\ 10^{2.5(\Omega + 0.5)} & ,-1.3 \le \Omega \le -0.5 \\ 1 & ,-0.5 \le \Omega \le 0.5 \\ 10^{-1.0(\Omega - 0.5)} & ,0.5 \le \Omega \le 2.5 \\ 0 & ,2.5 < \Omega \end{cases} \quad (74)$$

$$\theta(\Omega_i) = \sum_{i=-1.3}^{2.5} P(\Omega - \Omega_i) \cdot \phi(\Omega) \quad (75)$$

This convolution reduces the spectral resolution. The smoothed bark scale spectrum is down-sampled by resampling every 1 Bark(0 - 5 kHz maps to 0 - 16.9 Bark).

9.2. LOUDNESS PREEMPHASIS

Equal loudness preemphasis is needed to compensate for the non-equal perception of loudness at different frequencies. It is done by equal-loudness curve

$$E(w) = \frac{(w^2 + k_1)w^4}{(w^2 + k_2)^2(w^2 + k_3)} \quad (76)$$

where $k_1 = 56.8 \times 10^6$, $k_2 = 6.3 \times 10^6$, and $k_3 = 0.38 \times 10^9$.

9.3. INTENSITY-LOUDNESS LAW

Perceived loudness, L(w), is approximately the cube root of intensity, I(w). Therefore this pre-emphasized function is then amplitude compressed using cubic root amplitude compression.

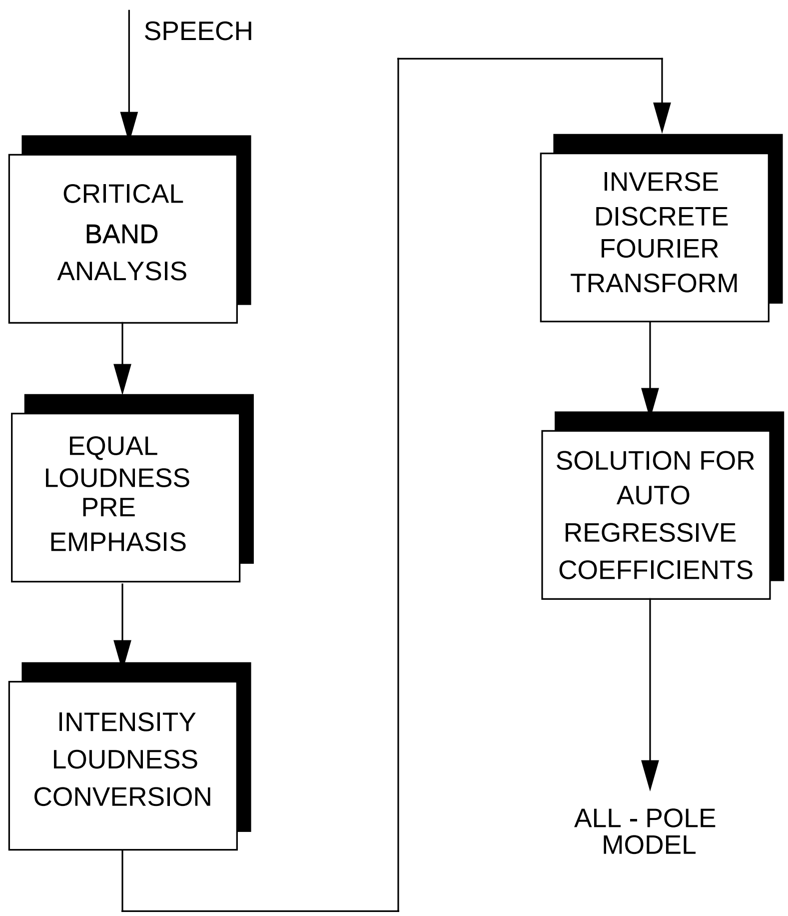

Figure 3. Block diagram of Perceptual Linear Predicton model

9.4. AUTOREGRESSIVE MODELLING

Final function is approximated by the spectrum of an all-pole model using the autocorrelation method of all-pole spectrum modeling[12]. The principle is to apply the inverse discrete Fourier transform (IDFT) and find the dual of its autocorrelation function.

The PLP-derived spectrum is more robust to noise compared to the LP-derived spectrum.

10. POLE-ZERO CELP

Pole-zero code excited linear prediction[13] is a new approach towards speech coding. A short-time spectral envelop of the speech the speech wave is modelled with a pole-zero filter. Advantage is that the zeros occurring in speech system, e. g. during nasalized speech sounds, can be modelled more accurately than with a traditional all-pole linear prediction filter. Knowledge of the excitation in the pole-zero CELP[18] coder leads to a linear but sub-optimum solution for the filter coefficients. It shows improved spectral matching in regions of anti-resonant nasalized sounds. Also the high frequency portion of the speech signal is matched better.

The speech sample *s(n)* is predicted from *p* immediately preceding speech samples and *q* immediately preceeding samples of the residual *e(n)*, hence:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + \sum_{k=1}^{q} b_k e(n-k) \quad (77)$$

In pole-zero CELP, an optimum all-pole filter *1/A(z)* and the coefficients of a closed loop pitch predictor *1/ P(z)* first are calculated. Then an all-zero filter *B(z)* is used in the synthesis of the speech as the pitch predictor *1/P(z)* and the all-pole filter *1/A(z)*. *B(z)* models the zero information in the speech that can not be modelled by either the codebook or the all-pole filter *1/A(z)*. The block diagram of the coder is given in Figure 4.

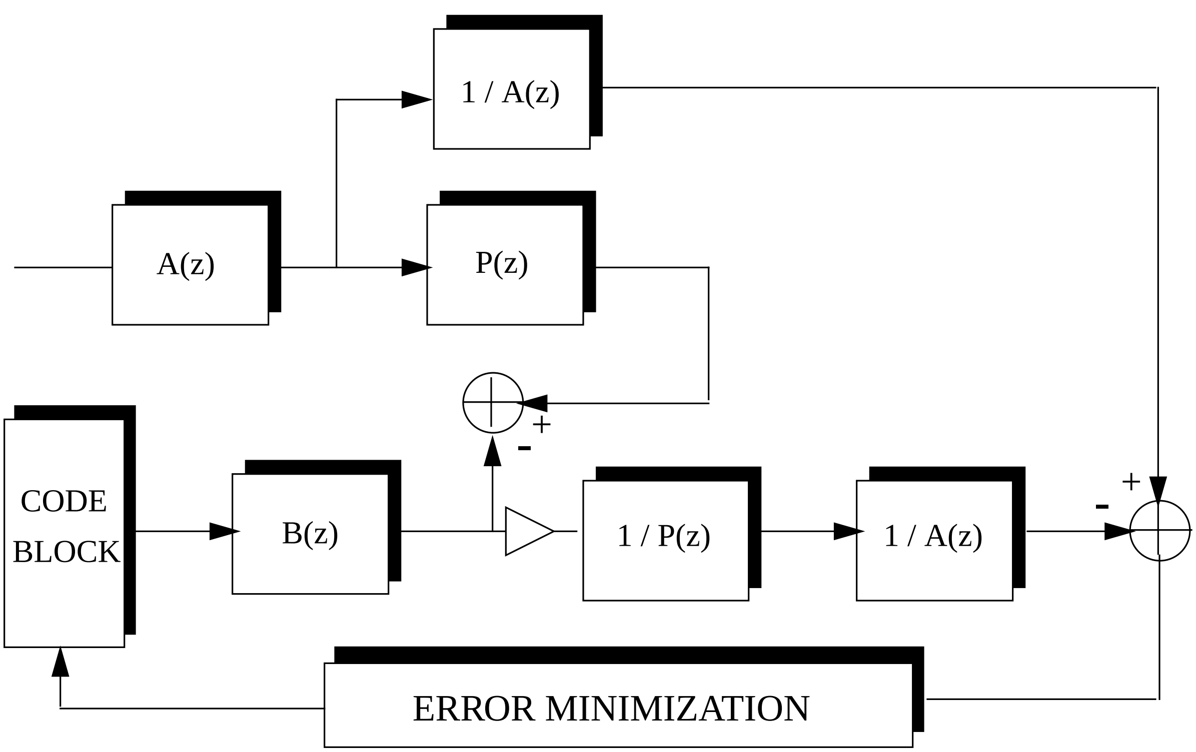

Figure 4. Pole-zero CELP coder where B(z) preceeds 1/A(z)

Objective and subjective comparisons were made between the pole-zero CELP coders and CELP coder. A 35 second segment of male speech, low-pass filtered to 4 kHz was used for the tests. Comparisons were based on segmental SNRs, power spectra and informal subjective listening tests. Segmental SNRs were calculated and they are shown in table 1.

Coder	Seg SNR dB
Standard CELP	14.96
Pole-zero CELP (B(z) pre-ceeds 1/A(z))	14.72
Pole-zero CELP (B(z) follows 1/A(z))	17.71

Table 1. Segmental SNR (dB) of different CELP coders

Thus it can be seen that for the pole-zero coder where the all-zero filter *B(z)* follows the all-pole filter *1/A(z)*, there is an increase of 2.7 dB in segmental SNR over the other two coders. However the subjective quality slightly becomes disimproved and spectral matching becomes better than traditional CELP. For the other pole-zero CELP, segmental SNR is very much similar to standard CELP but the spectral matching improves, especially for nasal sounds.

11. SUMMARY

In this paper we have attempted to cover the linear prediction. The emphasis is given to linear prediction as applied to speech. A number of properties and extensions of the basic linear prediction mathematics have been discussed, including speech synthesis structure, formant estimation, and spectral analysis. Using linear prediction we are able to provide an all-pole filter to model the vocal tract. LP analysis is also used for speaker identification[15]. But this model is not satisfactory to model the zeros manifested by antiresonances in the spectrum of speech sounds. A more generalized approach is to use pole-zero CELP which gives get better spectral matching for nasalized sounds. Similar approaches may be used to perform linear interpolation. In this technique least squares minimization technique can be used to estimate unknown data values in terms of past and future values. Such approaches may lead to more accurate pitch extraction in interpolating between data samples and may have applications in understanding non-minimum phase characteristics of speech.

REFERENCES

[1] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Engelwood Cliffs, New Jersy, USA, 1978.

[2] J.D. Markel and A.H. Gray Jr., *Linear Prediction of Speech*, Springer-Verlag, Berlin Heidelberg, New York, USA, 1976.

[3] R. J. Duncan, V. Mantha, Y. Wu and J. Zhao, “Implementation and Analysis of Speech Recognition Front-Ends,” http://www.isip.msstate.edu/resources/ece_4773/projects/1998/group_signal, *Institute for Signal and Information Processing*, Mississippi State University, USA, November 1998.

[4] J. Picone, “Signal Modelling Techniques in Speech Recognition,” *IEEE Proceedings*, vol. 81, no. 9, pp. 1215-1247, September 1993.

[5] V.R. Vishwanathan and J. Makhoul, “Quantization Properties of Transmission Parameters in Linear Predictive Systems,”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 3, pp. 309-321, June 1975.

[6] T. Robinson, “Speech Analysis,” http://svr-www.eng.cam.ac.uk/~ajr/SA95/Speech Analysis.html, University of Cambridge, U.K., October 1999.

[7] J. Makhoul, and J. Wolf, “Linear Prediction and Spectral Analysis of Speech,” BBN Report No. 2304, August 1972.

[8] B.S. Atal and S.L. Hanauer, “Speech Analysis and Synthesis by Linear Prediction of the speech wave,” *J. Acoust. Soc. Am.*, Vol 50, pp. 637-655, 1971.

[9] H. Hermansky, “Perceptual Linear Predictive (PLP) Analysis of Speech.” *Journal of the Acoustical Society of America*, vol. 4, pp.

1738-1752, 1990.

[10] L.R. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1993.

[11] J. Proakis and D.G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1996.

[12] J. Makhoul, "Spectral Linear Prediction: Properties and Applications," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-23, pp. 283-296, 1975.

[13] M. Dunn, B. Murray andA.D. Fagan, "Pole-Zero Code Excited Linear Prediction Using a Perceptually Weighted Error Criterion," *ICASSP-92*, vol.1, pp. 637-639, 1992.

[14] B.S. Atal, "Predictive Coding of Speech at Low Bit Rates," *IEEE Transactions on Communications*, vol. 30, no. 4, pp. 600-614, April 1992.

[15] B.S. Atal, "Linear Prediction For Speaker Identification," *Journal of the Acoustical Society of America*, vol. 55, no. 6, pp. 1304-1311, June 1974.

[16] S.L. Marple, Jr., *Digital Spectral Analysis With Applications*, Prentice-Hall, Engelwood Cliffs, New Jersy, USA, 1987.

[17] L.R. Rabiner, B.S. Atal and M.R. Sambur, "LPC Prediction Error Analysis of Its Variation with the Position of the Analysis Frame," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-25, no. 5, pp. 434-442, October 1977.

[18] J.D. Markel and A.H. Gray Jr., "A Linear Prediction Vocoder Simulation Based Upon the Autocorrelation Method," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-22, no. 2, pp. 124-134, April 1974.

[19] E.M. Hofstetter, "An Introduction to the Mathematics of Linear Predictive Filtering as Applied to Speech Analysis and Synthesis," *Technical Note* 1973-36, MIT Lincoln Labs, July 1973.

[20] S.S. McCandless, "An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-22, no. 2, pp. 135-141, April 1974.