Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (Toronto, ON), pp. 2405-2408, May 1991

M3.12

BACKWARD ADAPTIVE PREDICTION: HIGH-ORDER PREDICTORS AND FORMANT-PITCH CONFIGURATIONS

Majid Foodeei[1] and Peter Kabal[1,2]

[1] Electrical Engineering
McGill University
Montreal, Quebec, H3A 2A7

[2] INRS-Télécommunications
Université du Québec
Verdun, Quebec, H3E 1H6

Abstract

Backward adaptive linear prediction is used in low-delay speech coders. A good redundancy removal scheme must consider both near-sample (formant) and far-sample (pitch) correlations. Two approaches are considered; (1) separate pitch and formant predictors and (2) a single high-order predictor This paper presents analysis and simulation results comparing the performance of several types of high-order backward adaptive predictors with orders up to 100 Issues in high-order LPC analysis, such as analysis methods, windowing, ill-conditioning, quantization noise effects, and computational complexity are studied The performance of the various analysis methods is compared with the conventional sequential formant-pitch predictor The Auto-correlation method (50-th order) shows performance advantages over the sequential formant-pitch configurations. Several new backward high-order methods using covariance analysis and a lattice formulation show much better prediction gains than the Auto-correlation method

1. Introduction

Predictors are characterized by an analysis frame which is used to adapt the coefficients The predictor is then applied to a block of samples If the analysis frame precedes the block, the predictor is said to be backward adapted No buffering of "future" samples is needed in backward adapted systems, allowing for lower processing delay Backward adaptation is also used when there is no explicit transmission of the predictor coefficient values Both the coder and decoder can use backward adaptation to update the coefficients from the reconstructed signal

Traditionally two filters, the formant predictor and the pitch predictor, are used to remove near and far-sample redundancies Consider network-quality speech coding with low-delay and no explicit transmission of side information over channels with errors. For such coders (e.g low-delay tree and CELP coders), effective low-delay coding is achieved through removal of inter-sample redundancy using backward adaptive prediction. The pitch filter uses a backward adaptive pitch lag and coefficient values. The formant filter uses backward adaptive coefficient values Erroneous lag estimates at the receiver can cause severe error propagation. Higher processing power has made the alternative of a single high-order predictor attractive In this configuration, the combined pitch and formant taps have fixed positions This type of predictor performs better in the presence of channel errors. For high-order predictors, we consider issues and problems such as ill-conditioning, windowing, complexity and performance.

2. Formant and pitch prediction

For separate formant and pitch predictors, the transfer functions are

$$F(z) = \sum_{i=1}^{N_f} a_i z^{-i}, \quad \text{and} \quad P(z) = \sum_{i=1}^{N_p} b_i z^{-M_p-i-1}, \qquad (1)$$

where N_f and N_p are the number of formant and pitch predictor coefficients (e g $N_f = 10$ and $N_p = 3$) and M_p is the pitch lag Note that M_p is updated along with the coefficients The adaptation of pitch parameters for the pitch predictor in the cascaded pitch/formant configuration is described in detail in Ref [1] The multi-tap pitch filter $P(z)$ allows for non-integer pitch lag estimation In a speech coder, a closed-loop configuration can be used to achieve frequency shaping of the quantization noise.

As an alternative, a single high-order predictor can be used to remove both near and far-sample redundancies [2] The order P of this predictor should be high enough to include the effects of pitch correlations

The adaptation of the prediction coefficients is done in a backward fashion. The general analysis method of Fig 1 may be used to represent the windowing of data and/or error to estimate the prediction coefficients using the least-squares methods [1] The speech input, $s(n)$, is multiplied by the data window $w_d(n)$ to give windowed speech signal $s_w(n)$, while multiplication of the error signal by the error window $w_e(n)$ results in the windowed error signal $e_w(n)$ The error or data window shape may have be rectangular, Hamming, exponential, etc. Barnwell [3] and others [4] have studied and proposed a class of filters obtained using the impulse response of casual pole/zero filters which provide easy control over the shape of the window (exponential windows belong to this class). The shape of such windows is suitable for the backward adaptation since there can be more emphasis on recent data

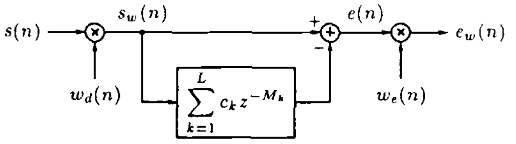

Fig. 1 Data window and error window

Covariance and windowed covariance methods

In the covariance method, windowing is performed only on the error signal ($w_d(n) = 1$ for all n) For a non-rectangular error window, the resulted method is called Windowed Covariance [5]. Tapered windows given smooth coefficient changes as the window is moved Minimization of the sum of windowed errors $E = \sum_{n=-\infty}^{\infty} e_w{}^2(n)$ results in the linear equations $\mathbf{\Phi a} = \mathbf{\Psi}$ ($\mathbf{\Phi}$ is the symmetric covariance matrix with components $\phi(i,j)$, $i,j = 1,2,\ldots,P$) This system can be solved using Cholesky decomposition to obtain the prediction coefficients Note that for the covariance methods, the number of terms entering into the correlation estimates is the same for all lags and is equal to the window length. However, correlations with large lags reach farther back into past data.

Modified Covariance method

The Modified Covariance method [6] is based on residual energy ratios It guarantees that the predictor is minimum phase

Auto-correlation method

If the windowing is done on the speech samples ($w_e(n) = 1$ for all n), the Auto-correlation method results. Since the auto-correlation matrix $\mathbf{R}$ ($\mathbf{Ra} = \boldsymbol{\alpha}$ replacing $\mathbf{\Phi a} = \mathbf{\Psi}$) is Toeplitz, the resulted linear system of equations can be efficiently solved using the Levinson recursion In calculating the auto-correlation components $r(i), i = 1,2,\ldots,P$, speech data outside the data window $w_d(n)$ are assumed

CH2977-7/91/0000-2405 $1.00 © 1991 IEEE

to be zero Fewer samples values are used for computing the larger lag values, making the Auto-correlation method inappropriate when the analysis window length N is comparable to the predictor order P Furthermore, tapered windows deemphasize the terms corresponding to large lags. Such windows can affect the numerical conditioning of the auto-correlation matrix by deemphasizing off-diagonal terms

Lattice and Covariance-Lattice methods

Lattice methods use a sequential solution formulation approach through which error minimization is done stage by stage. These methods do not assume optimality of the previous stages (in the Auto-correlation method, computational savings are due to this assumption) In these methods, a distinction is made between forward and backward errors At each stage m (representing a m-pole model) the forward and backward error (residual) signals, $f_m(n)$ and $b_m(n)$, are defined. The input speech is $s(n)$, $r(n)$ is the final prediction error (residual) signal, and n is the time index The filtering action of the lattice at stage m is described by

$$f_{m+1}(n) = f_m(n) - K_{m+1}(n) b_m(n-1) \quad \text{and} \tag{2.a}$$

$$b_{m+1}(n) = -K_{m+1}(n) f_m(n) + b_m(n-1), \tag{2.b}$$

where the K_m's are the reflection coefficients The initial forward and backward error signals are set to $s(n)$. One may minimize a combination of the forward and backward error energies. The minimization of the error energies can be expressed in terms of the following quantities

$$\begin{aligned} F_m(n) &= < f_m^2(n) > \\ B_m(n) &= < b_m^2(n) > \\ C_m(n) &= < f_m(n) b_m(n) > \end{aligned} \tag{3}$$

where the operation $<\ >$ is either an expectation or an appropriate time average. The Burg algorithm minimizes a combination of the forward and backward error energies with an error window $w(n)$,

$$E_m(n) = \sum_{k=-\infty}^{n} w(n-k) \frac{[f_m^2(k) + b_m^2(k)]}{2} \tag{4}$$

The choice of weighting factor 1/2 guarantees minimum phase [4] If an exponential window is used in the above formulation, the Exponential Window Lattice method [4] results The minimization of the above weighted error with respect to the $K_m(n)$ leads to the update formula

$$\begin{aligned} K_{m+1}(n+1) &= \frac{2C_m(n)}{F_m(n) + B_m(n)}, \quad \text{where} \\ C_{m+1}(n) &= \sum_{k=-\infty}^{n} w(n-k) f_m(k) b_m(k-1) \quad \text{and} \\ F_{m+1}(n) + B_{m+1}(n) &= \sum_{k=-\infty}^{n} w(n-k)[f_{m-1}^2(k) + b_{m-1}^2(k-1)] \end{aligned} \tag{5}$$

The Covariance-Lattice method [4, 7] uses the recursion formulations of the regular Lattice (Burg) method to obtain a more computationally efficient procedure The expected forward, backward, and cross error energies (Eqs 3) and the resulting update equation for the reflection coefficients are rewritten in terms of $\phi(k, i)$, the covariance of the signal The reflection coefficients are updated using the estimated covariance (Makoul Covariance-Lattice method [4])

An important modification to the above calculation was suggested by Cumani [7] To better fit the method to fixed-point arithmetic, the quantities are scaled for better numerical stability The price is a slight increase in computational complexity (Table 1) Our later results, indicate that this method has excellent numerical properties in the context of backward adaptive high-order predictors. Strobach [8] discusses this method further and generalizes it to the methods of *Generalized Residual Energy* (GRE) with the same numerical properties In this general class (which also includes the solution of the true Recursive Least-Squares (RLS) Covariance-Lattice estimation problem), coefficients are constructed completely anew at each step, avoiding round-off error accumulation As seen in Table 1, the computational complexity disadvantage of the Cumani algorithm ($O(P^3)$) can be overcome using Strobach Covariance-Lattice methods ($O(P^2)$) The class of

Method	Computation order *	$P = 10$	$P = 50$
Auto-correlation	$(PN+P^2)/M$	1700	10500
Covariance	$(PN+P^3/6+3P^2/2)/M$	1917	32583
Lattice (Burg)	$5PN$	8000	40000
Cov.-Lat (Makhoul)	$(PN+P^3/2+2P^2)/M$	2300	75500
Cov.-Lat (Cumani)	$(PN+4P^3/3-P^2)/M$	2833	172167
GRE (Strobach)	$(PN+3P^2)/M$	1900	15500

* First order terms have been neglected

Table 1 Comparison of the computation costs ($N = 160$)

such algorithms is also termed *Pure Order Recursive Ladder Algorithms (PORLA)*. Note that the adaptation of parameters in the Burg Lattice (and Exponential Window Lattice) is performed on a sample-by-sample basis while the Makhoul and Cumani Covariance-Lattice methods also allow for less frequent updates

3. Issues in high-order predictors

Consider the prediction error filter $A(z) = 1 - F(z)$, with the LPC analysis done in a backward fashion using the clean unquantized signal. A quantization model is introduced later. Even though the update rate for the LPC coefficients is not limited by transmission rate considerations in a backward adaptive configuration, for computational complexity considerations, we use less frequent update rates The effects of quantization on prediction gain can also be seen in a full coder in a companion paper [9]

Ill-conditioning

Due to the low-pass filter before the Analog-to-Digital-Converter (ADC), artificially low eigenvalues for the Covariance matrix Φ are produced (Covariance and Auto-correlation methods) These eigenvalues are related to the missing high frequency components in the sampled speech signal near half the sampling frequency This condition creates an almost singular covariance matrix which results in non-unique solution for the prediction coefficients The small eigenvalues produce artificially high prediction coefficients which if used, can cause problems [6] We term this "physical ill-conditioning".

An almost singular matrix is sensitive to round-off errors. As well, in high-order analysis, "numerical ill-conditioning" can exacerbate the round-off errors due to finite precision The high frequency correction procedure described in [6] decreases the physical ill-conditioning problem by using a new covariance matrix which is obtained by adding another matrix proportional to the covariance matrix of the high-pass filtered white noise to the original covariance matrix We use a simpler white noise correction (adding a small correction only to the diagonal elements). The effect of white noise correction scheme is reduction in the dynamic range of signal spectrum. We found that the white noise correction was preferable to high-pass correction Although the results presented are after the ill-conditioning "cure", the numerical error sensitivity remains an important issue. In the simulations, if a singular or almost singular matrix was encountered, the previous set of prediction coefficients was reused As the number of ill-conditioning cases increases in number, the effect is reflected in a decline in the prediction gain curve.

Window considerations

Window shape and size affect other analysis issues such as formant/pitch "capture", ill-conditioning and complexity An analysis window size and shape has to be carefully chosen in the case of single high-order predictor since it has to be suitable for both formant and pitch correlation lags. This choice is made even more difficult in the case of the Auto-correlation method due to the block edge effects In the cascaded formant-pitch configuration, the analysis frames of the near and far-sample predictors can have different size and update rate. Semi-infinite pole/zero windows (e g Barnwell windows and multiple exponential windows) are useful for backward adaptive processing. For example, Auto-correlation analysis using a 2 pole Barnwell (cascaded filters were used as recommended by [2]) outperforms the Hamming window (best length) by about 2 dB for female utterances (improvement for males was smaller) Error windowing using pole/zero windows is useful in the case of lattice analysis. Pole/zero windows also have computational advantages since they can be implemented recursively. As found in our experiments, the lack of taper at one end results in worse numerical problems than for windows such as the Hamming window which are tapered at both ends. Slightly longer effective lengths (see [3] for discussion on effective length of pole/zero windows) produce better results for higher order windows However, it was found that windows with very long "tails" resulted in poor performance

The prediction methods are based on an estimate of the correlation. The length of the analysis window must be large enough to provide a valid estimate of the correlations. The minimum formant frequency is around 270 Hz (males) Assuming a sampling frequency of 8000 Hz, this corresponds to a near-sample correlation lag of 30 samples A suitable formant analysis frame is 40–80 samples The range of pitch frequencies which has to be considered for natural speech is from 64 Hz to 400 Hz Typical male speech has an F_0 (fundamental frequency) range of 80 to 160 Hz With the average male and female F_0 at 132 and 223 Hz, the corresponding distances between pitch pulses in the time domain signal are 61 and 36 samples Hence from maximum and average pitch lag considerations, correlations corresponding to 120 and 61 lags (maximum and average) are needed. For the far-sample correlations of lowest pitch (64 Hz corresponding to 125 samples), a suitable effective length for the analysis frame (Auto-correlation method) is around 350 samples. If an analysis frame of such size is used, the stationary assumption for the formant characteristics may no longer be valid and formant variations may not be tracked faithfully Hence there is conflict between the desired formant and pitch analysis frame sizes (especially for the Auto-correlation method) and a compromise has to be made Traditionally an analysis frame of 120–160 samples (15–20 ms) have been used for formant analysis The experiments of this work indicated that although for higher order predictors a slightly longer analysis frame size results in a better performance, the frame sizes can still be used for Auto-correlation method high-order predictors. However, for the Covariance-Lattice methods, the best window length was around 100 samples.

4. Comparison of methods

Analysis methods

The order of the predictor should be high enough to include past samples as far back as the one corresponding to the lowest possible pitch (corresponding to order 125). With such a high order, the conflict between the desired formant and pitch analysis frame size may not allow for the full exploitation of the pitch and formant redundancies in a single predictor. In the previously published graphs of high-order predictor gain versus the prediction order, the curve becomes flat after order 50 [2]. This conflict is not as much of a problem for the Covariance/Covariance-Lattice methods as it is for the Auto-correlation method

The Covariance method showed severe ill-conditioning for high orders. So much so that the prediction gains dropped below acceptable levels White noise correction "cured" the ill-conditioning to some degree Fig. 2 shows a comparison of several Covariance techniques (160 sample blocks, updated every 20 samples). The curves labelled Covariance are with white noise correction. The Modified Covariance method has better numerical properties and produced better results, even though ill-conditioning still persists at high orders. The figure also shows the performance of the Exponentially Windowed (double exponential) Lattice method. The dropoff in performance for higher orders is due to numerical problems.

The best results were obtained when Covariance-Lattice (Cumani) was used. This is shown in Fig. 3. Most of the advantage comes from the better numerical properties of this method Windowing of the error can further improve the conditioning of the Covariance methods A Hamming window was applied to the Modified Covariance method, with results shown in the same figure The Covariance curves are not strictly monotonic for high orders. This can be attributed to two effects: remaining ill-conditioning and the fact that in backward adaptation,

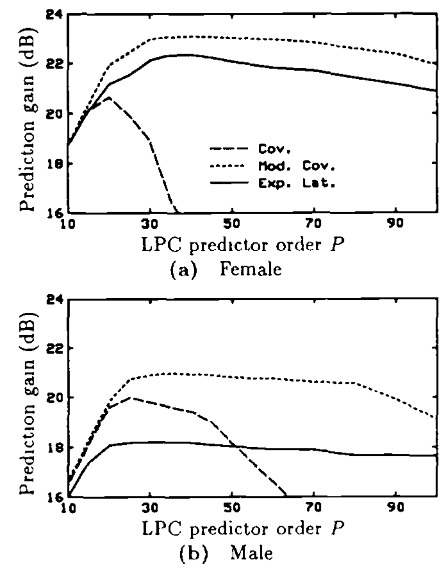

Fig. 2 Comparison of prediction gains showing the effect of ill-conditioning

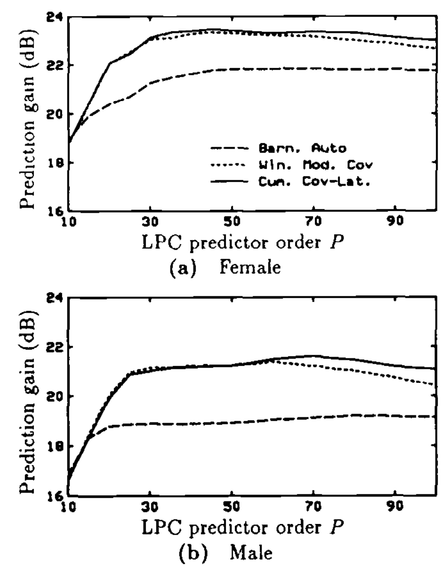

Fig. 3 Comparison of prediction gains for different methods

the data for the correlation calculation is "stale" and may not reflect the current statistics, especially for large lags

Fig. 3 also shows the results for the Auto-correlation method (Barnwell window). The Auto-correlation curves are monotonic but increase slowly if at all beyond order 50. Ill-conditioning is not a severe problem. The Auto-correlation curves lie well below the Covariance curves for larger orders, since the Auto-correlation method suffers from window shape effects.

Fig. 4 compares the segmental SNR for the Auto-correlation method, orders 10 and 50. Since the pitch lags for the female are less than 50, the order 50 analysis captures the female pitch; there is a substantial difference between the order 10 and order 50 results For the male (pitch around 70–80 samples), the pitch is out of reach for the order 50 predictor.

The computational cost comparison among various methods is shown in Table 1 The update rate M is an important factor when deciding among various alternatives Note that for the Lattice method $M = 1$ Experiment results investigating the update rate effect show

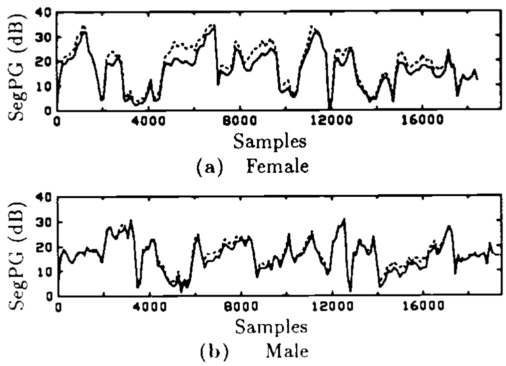

Fig. 4 Segmental prediction gain (segPG) comparison between speech analyzed using the LPC of order 10 (solid line) and of order 50 (dashed line) using Auto-Correlation method

that for the high performance methods, a less frequent update rate can be used since the performance degradation due to less frequent update is small

Backward adaptive effects and quantization noise

The experiments up to this point use backward adaptive prediction based on the clean speech signal For the purpose of evaluating the effect of quantization noise, a model assuming the following is used (1) The quantization noise is white, (2) The quantizer noise energy depends on the energy of the prediction residual (3) The quantizer SNR (prediction residual energy to quantization error energy) is fixed for an analysis frame (160 samples) The model is implemented by adding a noise energy to the diagonal elements of the Covariance matrix or the Auto-correlation matrix The level of noise is iteratively adjusted for each analysis interval until the specified SNR is achieved: (0) The initial quantization noise energy is set to a fraction of the input signal energy. (1) The correlation matrix is augmented with the quantization noise terms (2) The prediction coefficients are calculated (3) The residual energy is calculated (4) The quantization noise energy is reevaluated from the quantizer SNR (5) If the quantization noise energy has significantly changed, repeat the process at step (1). Table 2 shows the prediction gains for a quantizer SNR of 10 dB for the Autocorrelation method The effect of the quantization noise is to reduce the prediction gain Note also that the increase in going from order 10 to order 50 is reduced. Other experiments show that the effect of quantization noise is to close the gaps between the various analysis schemes (the ordering remains essentially the same as for clean speech) Note also that quantization noise has the same beneficial effect on ill-conditioning as the white noise correction.

Analysis based on	Gender	order 10 gain dB	order 30 gain dB	order 50 gain dB
1- Clean speech 2- Quantized speech	Female	18 93 18 51	21 25 20 27	21 83 20 83
1- Clean speech 2- Quantized speech	Male	16.84 16.74	18.89 18 51	18.91 18 17

Table 2 Comparison between analysis based on clean signal and simulated quantized signal for female and male utterances, Auto-correlation method

Comparisons with cascade formant-pitch configurations

Cascade formant-pitch configurations can have several forms: (1) Sequentially optimized, (2) Jointly optimized [1], and (3) Decoupled [10] These methods have the advantage of selecting independent analysis frame size for the formant and pitch predictors. Method (2) which produces the best overall prediction gain has a very high computational cost A problem with the cascade formant-pitch filters is their poor performance with channel errors; the lag estimate may be in error and cause severe error propagation Method (3) trades off performance for error robustness. For our comparisons (error propagation is not explicitly considered), we use the sequentially optimized formant/pitch configuration with a high-order predictor. During the unvoiced segments of the speech, the pitch predictor is turned off.

Table 3 compares the performance of the cascade filters with a single high-order predictor (Auto-correlation analysis, order 50), both types of predictors are updated every 5 samples. The results show that the high-order filter outperforms the F-P cascade for both the male and female speakers. Previous considerations indicated that the high-order filter does not capture the pitch of the male speaker However, the larger number of taps more than makes up for the loss (at least in terms of average prediction gain). Further experiments with a 50+3 F-P filter, show that for the female, the pitch part does not help (in fact, slightly degrades the prediction gain) However, for the male speaker, the pitch part gives a further gain (0 5 dB) which is perceptually important

method	Gender	formant gain dB	pitch gain dB	overall gain dB
F-P (10+3) Single high-order	female	16 5 –	1.2 –	17 7 19 9
F-P (10+3) Single high-order	male	14 8 –	0 7 –	15 5 17 0
F-P (50+3) Single high-order	female	19 9 –	−0 7 –	19 2 19 9
F-P (50+3) Single high-order	male	17 0 –	0 5 –	17.5 17 0

Table 3 Comparison of overall prediction gains between sequential formant-pitch configuration and high-order single predictor configuration

5. Summary and conclusions

Backward adaptation is used in low-delay speech coding applications. A single high-order filter provides excellent overall prediction gains. Compared to a cascade formant-pitch predictor, the high-order filter is more appropriate in the presence of channel errors. The Barnwell window Auto-correlation method with 50-th order predictor performs well for female utterances. Covariance methods promise even more prediction gain, since they can capture large lag pitch redundancies However, they are plagued with numerical problems The Cumani Covariance-Lattice with white noise correction overcomes these problems to a large degree. However, it is computationally intensive Further work is needed to find an appropriate compromise between complexity (varying the order and update rate) and performance. In addition, the PORLA algorithms (Strobach) might have a place in backward adaptive high-order predictors for speech coding.

References

1. P. Kabal and R P Ramachandran, "Joint Optimization of Linear Predictors in Speech Coders," *IEEE Trans. Acoust , Speech, Signal Processing*, vol ASSP-37, pp. 642–650, May 1989
2. J. H. Chen, "A Robust low-delay CELP speech coder at 16 kb/s," *IEEE workshop on speech coding for telecommunications*, Vancouver, Canada, Sept 1989
3 T P. Barnwell, III, "Recursive Windowing for Generating Autocorrelation Coefficients for LPC analysis," *IEEE Trans Acoust , Speech, Signal Processing*, pp. 1062–1066, Oct. 1981
4. J. I Makhoul and L. K. Cosell, "Adaptive Lattice Analysis of Speech," *IEEE Trans Acoust., Speech, Signal Processing*, vol ASSP-29, pp. 654–659, Apr 1981.
5. S. Singhal and B. S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates," in *Proc IEEE Int Conf Acoust , Speech, Signal Processing*, San Diego, CA, March 1984, pp 1 3.1–1.3.4.
6. B. S. Atal and M. R Schroeder, "Predictive Coding of Speech Signals and Subjective Error Criteria," *IEEE Trans Acoust , Speech, Signal Processing*, vol. ASSP-27, pp. 247–254, April 1979
7. A. Cumani, "On a Covariance Lattice Algorithm for Linear Prediction," in *Proc. Int. Conf Acoust., Speech, Signal Processing*, Paris, Frace, 1982, pp. 651–654
8. P. Strobach, "Pure Order Recursive Least-Square Ladder Algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol ASSP-34, Aug. 1986, pp. 880–897.
9 M Foodeei and P. Kabal "Low-Delay CELP and Tree Coders: Comparison and Performance Improvements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Ont. May 1991
10. V. Cuperman, A. Gersho, R. Pettigrew, J. J. Shynk and J. H. Yao "Backward Adaptation for Low Delay Vector Exitation Coding of Speech at 16 Kbits/s," *ICASSP-89* pp 34.2.1–34.2.5