LOW BIT RATE SPEECH CODING METHODS AND A NEW INTERFRAME DIFFERENTIAL CODING SCHEME FOR LINE SPECTRUM PAIRS

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND

ELECTRONICS ENGINEERING

AND THE INSTITUTE OF ENGINEERING AND SCIENCES

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Engin ERZIN

June 1992

LOW BIT RATE SPEECH CODING METHODS AND A NEW INTERFRAME DIFFERENTIAL CODING SCHEME FOR LINE SPECTRUM PAIRS

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND

ELECTRONICS ENGINEERING

AND THE INSTITUTE OF ENGINEERING AND SCIENCES

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF
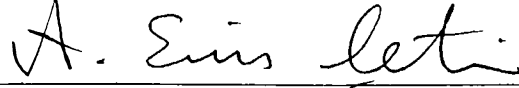MASTER OF SCIENCE

By
Engin Erzin
June 1992

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.
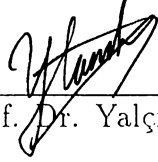
_____

Assoc. Prof. Dr. A. Enis Çetin(Principal Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

_____

Prof. Dr. Yalçın Tanık

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.
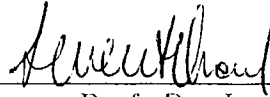
_____

Assoc. Prof. Dr. Levent Onural

Approved for the Institute of Engineering and Sciences:

_____

Prof. Dr. Mehmet Baray
Director of Institute of Engineering and Sciences

ABSTRACT

LOW BIT RATE SPEECH CODING METHODS AND A NEW INTERFRAME DIFFERENTIAL CODING SCHEME FOR LINE SPECTRUM PAIRS

Engin Erzin
M.S. in Electrical and Electronics Engineering
Supervisor: Assoc. Prof. Dr. A. Enis Çetin
June 1992

Low bit rate speech coding techniques and a new coding scheme for vocal tract parameters are presented. Linear prediction based voice coding techniques (linear predictive coding and code excited linear predictive coding) are examined and implemented. A new interframe differential coding scheme for line spectrum pairs is developed. The new scheme reduces the spectral distortion of the linear predictive filter while maintaining a high compression ratio.

Keywords : Speech coding, linear predictive coding, vocal tract parameters, pitch, code excited linear prediction, line spectrum pairs.

ÖZET

AZ BITLE SÖZ KODLAMA METODLARI VE DOĞRUSAL SPEKTRUM ÇİFTLERİ İÇİN YENİ BİR ÇERÇEVELERARASI FARK KODLAMA YAPISI

Engin Erzin
Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans
Tez Yöneticisi: Doç. Dr. A. Enis Çetin
Haziran 1992

Az bitle söz kodlama teknikleri irdelenmiş ve ses yolu parametreleri için yeni bir kodlama yapısı sunulmuştur. Bu amaçla çeşitli doğrusal öngörü kökenli söz kodlama teknikleri (doğrusal öngörülü kodlama ve kod beslemeli doğrusal öngörülü kodlama) incelemiş ve gerçeklenmiştir. Ayrıca Doğrusal Spektrum Çiftleri için yeni bir çerçevelerarası fark kodlama yapısı geliştirilmiştir. Önerilen bu yapının doğrusal öngörülü süzgeçin spektral bozulmasını azaltmada başarılı olduğu gösterilmiştir..

Anahtar kelimeler : Söz kodlaması, doğrusal öngörülü kodlama, ses üretim parametreleri, perde, kod beslemeli doğrusal öngörü, doğrusal spektrum çiftleri.

ACKNOWLEDGMENT

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Speech is an important tool of communication. Webster's Dictionary defines speech as "the power of audible expression, talk, oral expression or communication". Therefore a speech signal is an information bearing signal. In this thesis coding and transmission of speech signals are studied.

An efficient speech coding and transmission method must be based on understanding of how human beings produce it. The main organs that help to the production of speech are the *larynx*, which contains vocal cords, and the *vocal tract*, which is a tube leading from the larynx along the pharynx and then branching into the oral cavity leading to the lips and through the nasal cavity to the nostrils.

Acoustic energy in speech can be generated in two different ways. The first mechanism produces voiced excitation in the larynx. The vocal cords oscillate quasi-periodically at an average rate of 110 times per second for man and about twice of that for a woman. The resulting voiced speech include all vowels and many consonant sounds. The second mechanism produces acoustic energy in speech using the turbulence created by the tongue or lips. The generated sounds in this way (such as, 's' or 'sh') are said to be voiceless and they generally play a less important role in speech than voiced sounds.

Human hearing system senses the loudness of the sound by the log of acoustic energy rather than its linear value. Therefore, doubling the energy in a sound leads linear increase in loudness. The maximum sensitivity of our amplitude hearing is in the 1 to 2 kHz range. This sensitivity falls off below 100 Hz and above 5-10 kHz depending on the age.

In terms of communication, speech is a signal with a message or information.

1

A speech message is usually preserved in two ways,

(i) by the message content (type (i) coder), and

(ii) by retaining the speech waveform in a form that is convenient for transmission and storage (type (ii) coder).

Waveform representation of speech which can be considered as a type (i) coder consists of concatenation of elements called phonemes. A set of phonemes forms a basis for speech signal and phonemes differ with different languages. A way of coding speech waveform is concatenating phonemes and this costs about 100 bits/sec transmission rate [1]. Although this may be the lowest rate that can be achieved, the concatenation technique can not sense the rate of speaking, the loudness and the emotional content of the speech, etc.

The acoustic speech signal can be translated into electrical signal by a transducer. The relative bandwidth of this signal is about 4 kHz. In many telecommunication applications the analog signal is filtered by a lowpass filter with cutoff at 3.6 kHz. After lowpass filtering, this signal is sampled with a sampling rate of 8 kHz. Usually the A/D converter uses 12 bits per sample. The simplest type (ii) speech coder is the Pulse Code Modulator (PCM), which is just a non-uniform quantizer. The PCM method converts 12 bit samples to 8 bit μ-law or A-law coded samples at 8 kHz sampling rate and this corresponds to 64 kbit/sec transmission rate [1].

Today's technology lets us process the discrete-time speech signal in a very flexible manner. There are microprocessors called digital signal processors, with 60 ns instruction cycle, that is we are capable of using about 2100 instructions for one sample of speech signal. Because of this, in recent years many computationally intensive speech coding techniques have been developed [2].

In this thesis we consider two ways of representing the speech signal based on linear modelling of the vocal tract. The Linear Predictive Coding (LPC) [1] and Code Excited Linear Predictive Coding (CELP) [2] methods are implemented by using the TMS320C25 digital signal processor and *soundtool* software of SUN-Sparc workstations. The main contribution of this thesis is a new interframe differential coding scheme for vocal tract parameters, this new coding method is used with LPC and CELP coders. The LPC and CELP coders are examined with their basic properties in Chapter 2 and 3, respectively.
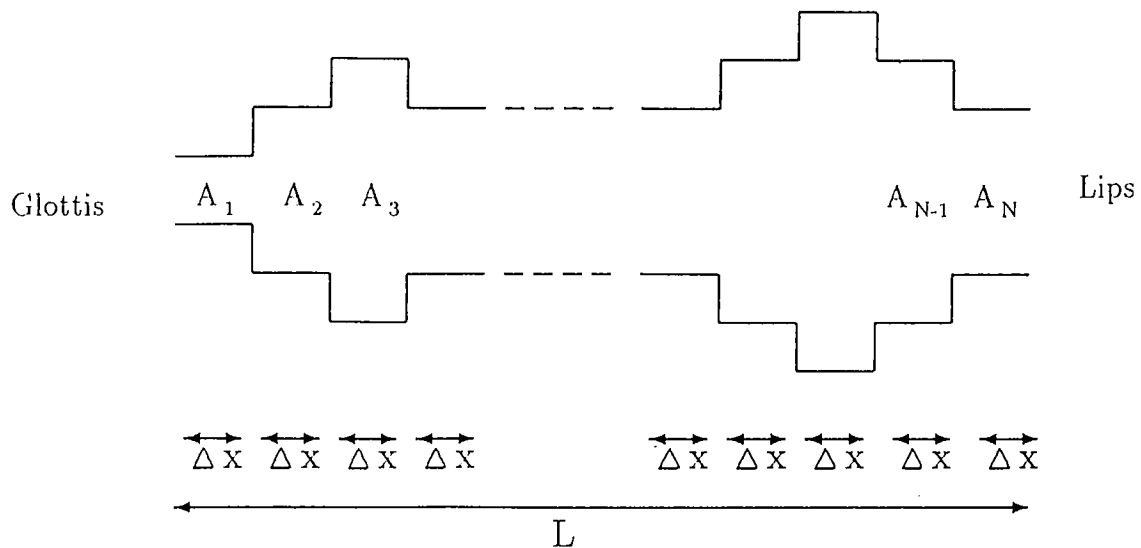
Figure 1.1: *Acoustic Tube Model of the Vocal Tract*

1.1 Linear Modelling of the Vocal Tract

In this section linear modelling of the vocal tract is described and parameters of this linear model called vocal tract parameters which are essentially important in many speech coding methods are defined.

The vocal tract plays an important role in speech generation. It is excited by the flow of air coming from the lungs. Although the vocal tract area function, $A(x, t)$, is considered to be a time-varying linear system, it is assumed to be constant for short time intervals (10-30 ms) during the speech generation process. Thus vocal tract is simulated by a concatenation of lossless acoustic tubes as shown in Figure 1.1. The length of each tube is $\Delta x = \frac{L}{N}$, where L is the overall length of the vocal tract and N is the number of the tubes. Wave propagating in this system can be represented [3] as in Figure 1.2 with the delays being equal to $\tau = \frac{\Delta x}{c}$ which is the time to propagate the length of one tube, where c is the velocity of sound. The parameters r_k are the reflection coefficients for the kth junction and given as,

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \tag{1.1}$$

where A_k is the area of the kth portion of the acoustic tube. This definition implies that $-1 \leq r_k \leq 1$ as all areas are positive.

For a sampling period of $T = 2\tau$, the equivalent discrete-time system for band-limited inputs can be obtained as shown in Figure 1.3-a. Since implementation of $\frac{1}{2}$ sample delays is not easy, a more desirable configuration with

whole delays can be achieved, Figure 1.3-b.
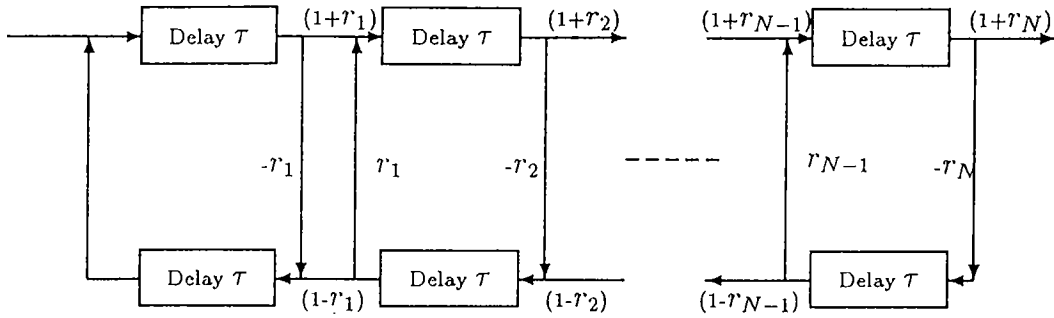
Figure 1.2: *Signal flow graph for lossless tube model of the vocal tract.*
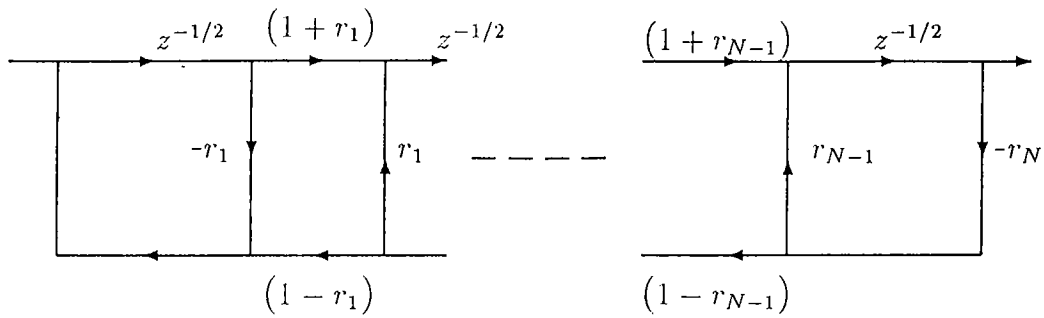
(a)
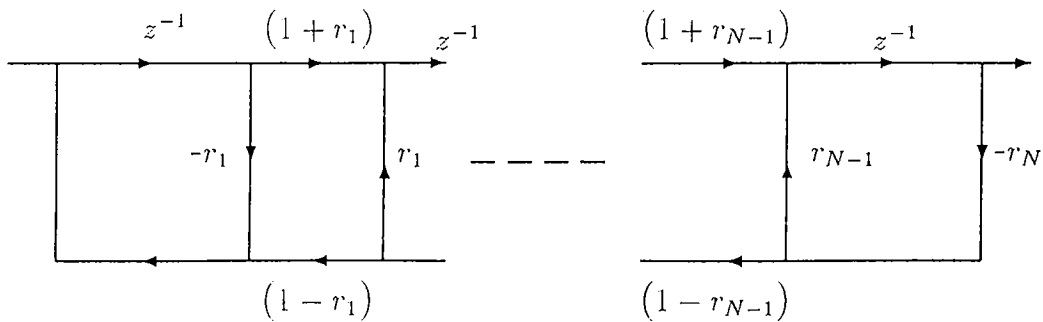
(b)

Figure 1.3: *(a) Equivalent discrete-time system for lossless tube model of the vocal tract, (b) equivalent discrete-time system using only whole delays.*

In general, the transfer function, $H(z)$, for a lossless tube model can be expressed as follows [1],

$$H(z) = \frac{\prod_{k=1}^{N}(1 + r_k)}{Y(z)} \qquad (1.2)$$

where $Y(z)$ can be obtained from the recursion: $Y_0(z) = 1$

$$Y_k(z) = Y_{k-1}(z) + r_k z^{-k} Y_{k-1}(z^{-1}), \quad k = 1, 2, ..., N.$$

$Y(z) = Y_N(z)$ that is $Y(z)$ will have the form,

$$Y(z) = 1 - \sum_{k=1}^{N} a_k z^{-k}.$$ (1.3)

Then the transfer function $H(z)$ reduces to an all pole linear filter form,

$$H(z) = \frac{G}{Y(z)}$$ (1.4)

where $G = \prod_{k=1}^{N}(1 + r_k)$.

The reflection coefficients, r_k's, are called the vocal tract parameters and most of the speech coding methods use these parameters to represent the linear predictive filter. Coding performance of the reflection coefficients is better than coding of the filter coefficients, as reflection coefficients lie in the range -1 to 1.

Finally, the lossless tube model produces an all pole linear filter, which works well in modeling the human speech production system.

Chapter 2

Linear Predictive Coding (LPC) of Speech

In this chapter we describe the linear predictive analysis and examine some necessary parameters for speech estimation and prediction, such as pitch value. Also, an implementation of the LPC vocoder (voice coder) is presented.

Linear predictive analysis method is a powerful speech analysis technique for estimating the basic speech parameters, such as pitch, vocal tract area function, etc [1]. Accurate estimation of the speech parameters and low computational complexity make this method a widely used one.

The basic idea of linear predictive coding is that the current speech sample can be estimated as the linear combination of past samples. Applying an error minimization criterion one can come up with a set of predictor coefficients for an all pole, linear, time-varying filter model. This all pole linear filter is excited either by quasi-periodic pulses (during voiced speech), or random noise (during unvoiced speech) to form the synthetic speech.

There are various ways of carrying out the linear predictive analysis of speech such as, the covariance method, the autocorrelation method, the lattice method, and the inverse filter formulation [1]. The most common ones are the autocorrelation and covariance methods. Although autocorrelation method always produces stable solutions, the performance of covariance method is slightly better than autocorrelation method especially for voiced speech [1]. In this chapter the covariance formulation is presented.

2.1 Covariance Method for Linear Predictive Analysis

The basic form of the LPC vocoder is given in Figure 2.1. In this case, the modelling of speech waveform is represented by a time-varying digital filter with steady-state system function:

$$H(z) = \frac{G}{A(z)} \tag{2.1}$$

where $A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}$.
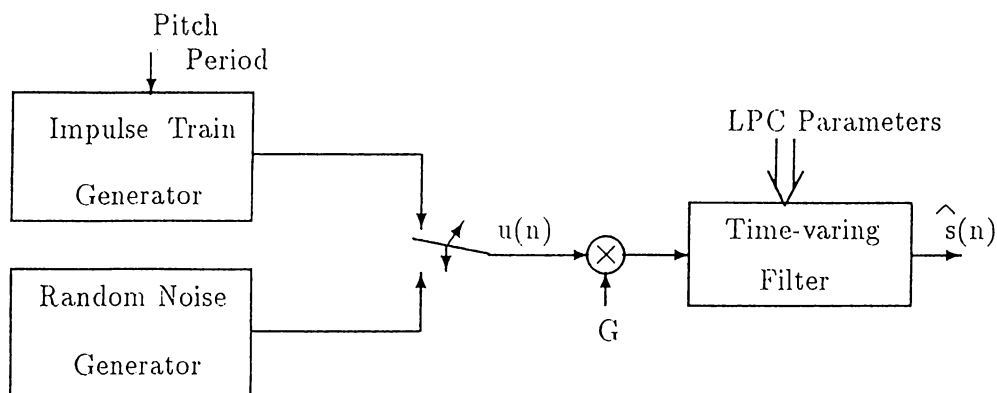
Figure 2.1: *LPC Vocoder Synthesizer*

LPC analysis is a frame-oriented technique which performs analysis for speech segments of duration 20-30 ms. The LPC system is excited by an impulse train for voiced speech and random noise sequence for unvoiced speech. Voiced/unvoiced decision and pitch period calculation are done in either time or frequency domain. A frequency domain method, cepstrum method, is examined in the following section. The other parameters used in the system are the gain, G, and the coefficients a_k of the LP filter. All these parameters slowly vary in time.

Let us define the prediction error,

$$e(n) = s(n) - \sum_{k=1}^{p} a_k s(n - k) \tag{2.2}$$

where $s(n)$ is the true value and $\sum_{k=1}^{p} a_k s(n - k)$ is the predicted value of the speech at time instant n. Then the short-time average prediction error is defined as,

$$E = \sum_{m} e^2(m) \tag{2.3}$$

$$= \sum_{m}(s(m) - \sum_{k=1}^{p} a_k s(m - k))^2 \tag{2.4}$$

where the range of the outer summation is determined according to the duration of speech segments which range from 20 to 30 ms.

In order to minimize the predictor error E, we can obtain a set of equations by setting $\frac{\partial E}{\partial a_k} = 0$ for $k = 1, ..., p$. After some algebraic manipulations we obtain the following equations:

$$\sum_m s(m - i)s(m) = \sum_{k=1}^{p} a_k \sum_m s(m - i)s(m - k), \quad i = 1, ..., p. \tag{2.5}$$

Let us define,

$$\phi(i, k) = \sum_m s(m - i)s(m - k) \tag{2.6}$$

then Eq. (2.3) reduces to

$$\sum_{k=1}^{p} a_k \phi(i, k) = \phi(i, 0), \quad i = 1, ..., p. \tag{2.7}$$

Equation (2.5) is a linear system of p equations with p unknowns, and it can be solved in an efficient manner [1]. In matrix form these equations become,

$$\begin{bmatrix} \phi(1,1) & \phi(1,2) & \phi(1,p) \\ \phi(2,1) & \phi(2,2) & \phi(2,p) \\ \vdots & \vdots & \vdots \\ \phi(p,1) & \phi(p,2) & \phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \phi(1,0) \\ \phi(2,0) \\ \vdots \\ \phi(p,0) \end{bmatrix} \tag{2.8}$$

and the pxp coefficient matrix is called the covariance matrix and Eq. 2.8 can be solved efficiently by Cholesky decomposition [4] or Levinson-Durbin recursion can also be used to determine the LPC parameters [5].

2.2 Voiced/Unvoiced Decision and Pitch Period Detection

In this section we examine the voiced/unvoiced characteristics of the speech signal. One important characteristic of speech is the periodic or nearly periodic nature of it, if it is voiced. This characteristic causes considerable redundancy which can be exploited by predicting the current samples from samples observed one period earlier. The number of glottal openings per second is closely associated with this periodic nature of speech segment and the repetition period is often called the pitch period. Estimation of the pitch period is an important problem in analyzing the speech waveform. Autocorrelation method [6], average magnitude difference function (AMDF) method [7], cepstrum method

[3],[8] are the widely used methods for estimating the pitch period. AMDF and autocorrelation methods have low complexity but weak performance on finding the true value of the pitch period compared to the cepstrum method. So the details of the cepstrum method is presented in this section.

Voiced/unvoiced decision and pitch period are the parameters which determine the excitation of the linear prediction filter in Figure 2.1. The energy and the periodicity of the speech signal are the factors for the voiced/unvoiced decision. Voiced speech has a periodic characteristics with a high RMS value, and unvoiced speech has a pseudo-random characteristics with a low RMS value. Voiced/unvoiced decision and pitch period detection can be made both in time or frequency domain. A frequency domain method, cepstrum method, is more reliable but it is more computationally complex than other time domain methods [3].

By considering the vocal tract model, the speech signal can be modelled as an output of the vocal tract excited by a vocal source as shown in Figure 2.2. Therefore the speech can be modeled as a convolution of a vocal source, $s(t)$,

$$\boxed{\text{Vocal Source}} \xrightarrow{\text{s(t)}} \boxed{\text{Vocal Tract} \quad \text{h(t)}} \xrightarrow{\text{f(t)} = \text{s(t)} * \text{h(t)}}$$
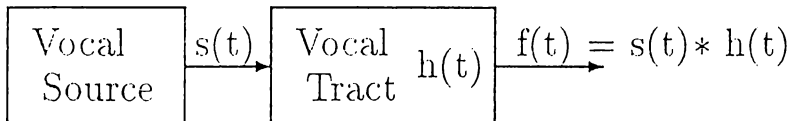
Figure 2.2: *Modelling of the speech production system*

with a vocal tract function, $h(t)$. In frequency domain this can be written as,

$$F(w) = S(w)H(w) \tag{2.9}$$

Then the autocorrelation function,

$$r(\tau) = \mathcal{F}^{-1}(|F(w)|^2) \tag{2.10}$$

expressed as a convolution of individual autocorrelation functions of $s(t)$ and $h(t)$ as,

$$r(\tau) = r_s(\tau) * r_h(\tau). \tag{2.11}$$

This convolution in some cases causes multiple peaks at the autocorrelation function $r(\tau)$ and voiced/unvoiced decision becomes difficult to make. One way of separating the effect of vocal tract from vocal source is the cepstrum method.

By computing the logarithm of both sides of (2.9) in frequency domain one can separate the two vocal function as follows,

$$Log|F(w)|^2 = Log(|S(w)|^2|H(w)|^2) \tag{2.12}$$

$$= Log|S(w)|^2 + Log|H(w)|^2 \tag{2.13}$$

and we define the cepstrum function by taking magnitude squares of the inverse Fourier transform of each sides,

$$C(\tau) = |\mathcal{F}^{-1}(Log|F(w)|^2)|^2 = |\mathcal{F}^{-1}(Log|S(w)|^2) + \mathcal{F}^{-1}(Log|H(w)|^2)|^2 \tag{2.14}$$

In this case the source and the tract effects become additive and the vocal tract part consists of low frequency components (in the order of seconds), vocal source part consists of high frequency components. If the speech segment is voiced then a peak corresponding to the source periodicity appears clearly in $C(\tau)$. An example is shown in Figure 2.3. If the frame is voiced (Figure 2.3.a) then a peak appears in the position of pitch period. If the frame is unvoiced (Figure 2.3.b) then we can not see a sharp peak, other than $\tau = 0$ location.

Figure 2.3: *Sample outcomes of cepstrum function for voiced (a) and unvoiced speech (b), respectively.*

Another cepstrum method is based on the computation of the cepstrum of the linear prediction error sequence (residual signal). The cepstrum of the residual sign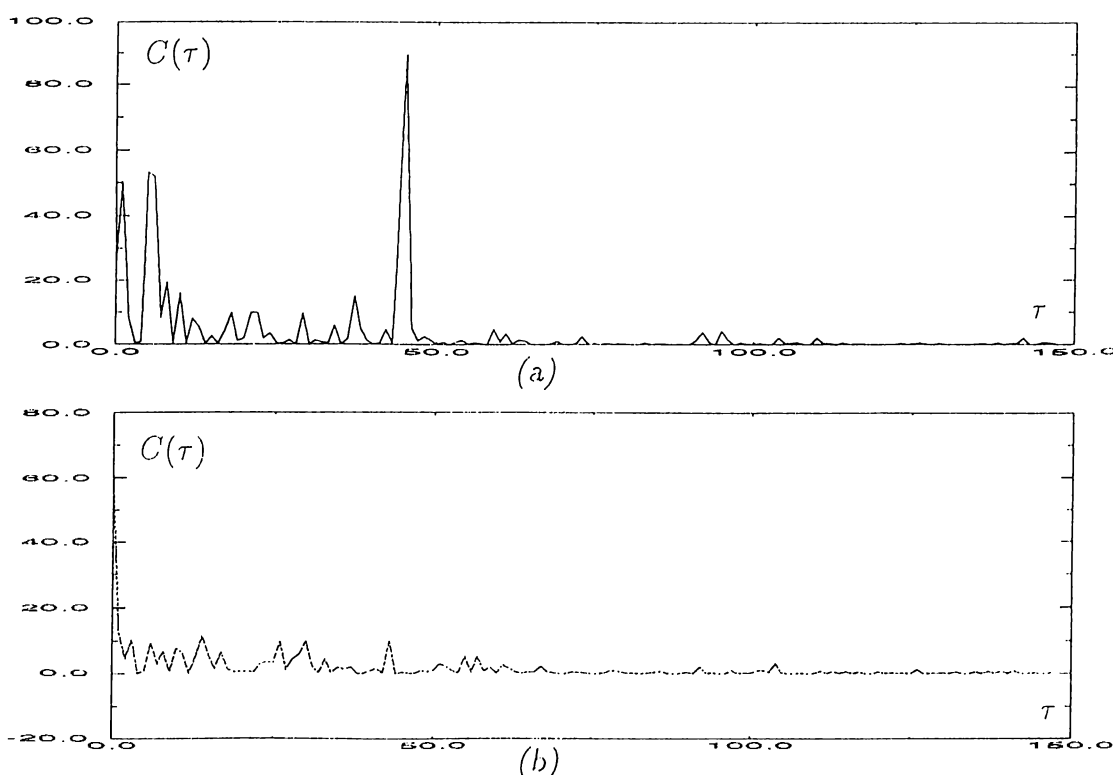al also shows a similar behavior as in Figure 2.3. The outcome of cepstrum function provides an easy voiced/unvoiced decision and gives a more reliable decision on the pitch value. We can also state that the cepstrum applied on residual signal gives better decisions [23].

2.3   Implementation of LPC Vocoder with TMS320C2X Micro-Processors

Real time implementation of the LPC vocoder system was realized as a group project in EE 526 DSP Laboratory course in 1990. The implementation was performed with TMS320C2X microprocessors. The system consists of three main blocks, first one is the A/D, D/A conversion of speech signal at 8 kHz, the second block consists of the analysis and synthesis of the speech signal and the third block consists of the serial communication link between analyzer and synthesizer as shown in Figure 2.4. Our LPC vocoder system is compatible with NATO standards [9], we use 10-th order linear prediction filter and a frame (180 samples) is represented by 54 bits. This corresponds to a transmission rate of 2.4 kbits/sec.

In our implementation we used TMS320C20 at the synthesizer part and TMS320C25 at the analyzer part. TMS320C20 and C25 micro-processors have 200 and 100 ns instruction cycles, respectively. Both of these processors are located on a plug-in PC-card which contains A/D and D/A converters and a serial transmission port.

2.3.1   Analysis

Analog speech signal is sampled at 8 kHz with a precision of 12 bits/sample. The discrete-time speech signal is processed in frames (each frame consists of 180 samples). After first order pre-emphasis $(1 - 0.9375z^{-1})$ voiced/unvoiced decision is made and then 10 predictor coefficients are determined by linear predictive analysis. Covariance formulation is used for linear predictive analysis. Based on H. Padır's M.Sc. thesis, the average magnitude difference method (AMDF) [7] is used for voiced/unvoiced decision. The AMDF method has a low complexity and its performance is acceptable. The AMDF method forms
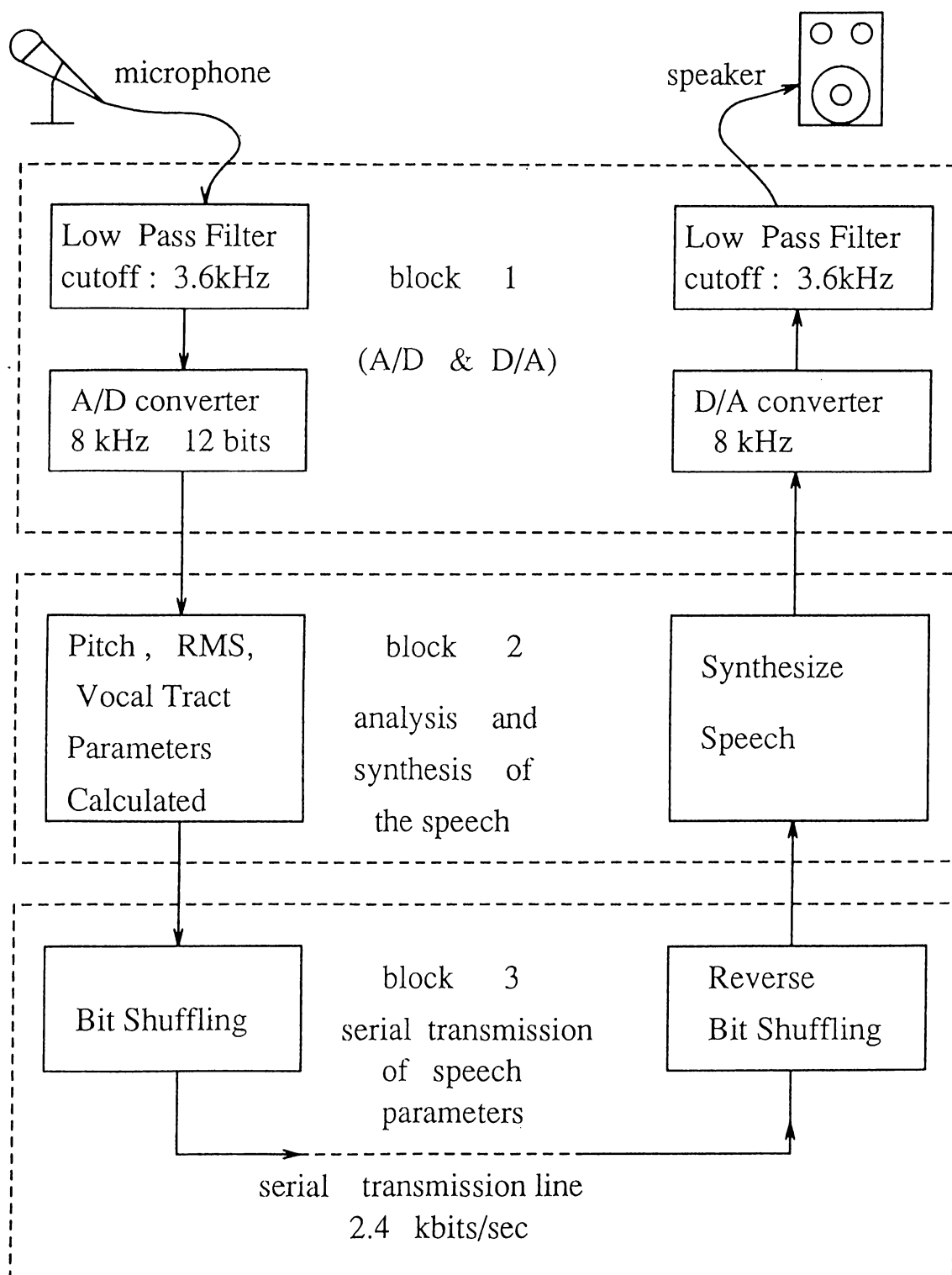
Figure 2.4: *LPC Vocoder Flow Diagram*

a difference signal between the delayed and the original speech and, at each delay, the absolute magnitude of the difference is evaluated. The difference signal is zero, if the delay is zero, and exhibits deep nulls at delays corresponding to the pitch period of voiced speech. The AMDF method is chosen for real time implementation because of its simple nature and it also gives a reasonable estimates of the pitch periods. In the analysis section the gain (RMS) of the frame is calculated, too. All of the estimated parameters are passed to a block which sends them through the serial port after a bit shuffling process. In our system the transmission is not synchronous, it is asynchronous, so that for each frame we have extra two start and stop bits which increase the transmission rate above the NATO standard rate, 2.4 kbits/sec. But the synchronous transmission can easily be implemented with a synchronization algorithm.

2.3.2 Synthesis

The synthesizer receives the shuffled bit stream and decodes the parameters of each frame from this stream. If the frame is voiced, then the input excitation sequence is a periodic impulse train. The periodicity of the impulse train is the estimated pitch value. Otherwise input excitation sequence is formed by using a pseudo-random number generating stream. Then the LPC filter whose coefficients are extracted from the coded bit stream is excited by either the impulse train or pseudo-white noise. In this way a synthetic speech signal is generated at the receiver. Finally, a de-emphasis ($\frac{1}{1-0.75z^{-1}}$) and RMS normalization is performed and an interrupt routine outputs the synthetic speech at a rate of 8 kHz.

2.3.3 Implementation of a LPC Vocoder on SUN-Sparc Stations

We also implemented an LPC vocoder on SUN-Sparc stations by using the *soundtool* software and the *C* compiler. In this implementation, we did not care about the computational complexity of the individual sub-algorithms of the LPC vocoder and selected algorithms to achieve the best performance. We compared some different algorithms for voiced/unvoiced decision and LPC parameters coding. The cepstrum method is performed for voiced/unvoiced decision both on speech signal and residual signal. Synthesized speech quality tests showed us that cepstrum method is better than AMDF method in

general and also cepstrum applied on residual has a better performance than cepstrum applied on speech signal. LPC parameters coding is performed by using Line Spectrum Pairs, which represent the linear predictive filter coefficients in a robust way. (Interframe differential coding of line spectrum pairs is described in Chapter 4). We also used a smoothing scheme for LPC parameters and pitch period. The smoothing removes discontinuities in the synthesized speech. Finally, we achieved a better quality synthesized speech than the real-time LPC-10 vocoder implementation described in the previous section. The developed software is available at the Electrical and Electronics Engineering Department Library.

Chapter 3

Code Excited Linear Prediction of Speech

In this chapter we present another speech coding method, Code Excited Linear Prediction (CELP) [2],[10],[11]. CELP coding is based on analysis-by-synthesis search procedures, vector quantization (VQ), and linear prediction (LP). The formant structure of the speech is modelled by a 10-th order LP filter. The long-term signal periodicity is modelled by an adaptive codebook, and the error from the linear prediction filter excited by codewords from the adaptive codebook is also vector quantized by using a fixed stochastic codebook. The optimal excitation vectors from adaptive and stochastic codebooks are selected by minimizing the error between synthesized and original speech in the M.S.E. sense.

For the 4.8 kbits/sec CELP, the stochastic codebook consists of 512 ternary valued (-1,0,+1) codewords, and adaptive codebook consists of 128 codewords which are refreshed by the previous excitation sequence at every subframe. 30 ms frame size is used at 8 kHz sampling rate and therefore each frame consists of 240 samples (or 4 subframes where each subframe consists of 60 samples). The transmitted CELP parameters are the stochastic and adaptive codebook indices and gains, and 10 line spectral pairs (LSP) as the vocal tract parameters.

3.1   Synthesis

The CELP synthesizer, shown in Figure 3.1, is both used in the receiver and the transmitter. The excitation is formed by stochastic and adaptive codebook vectors.

15
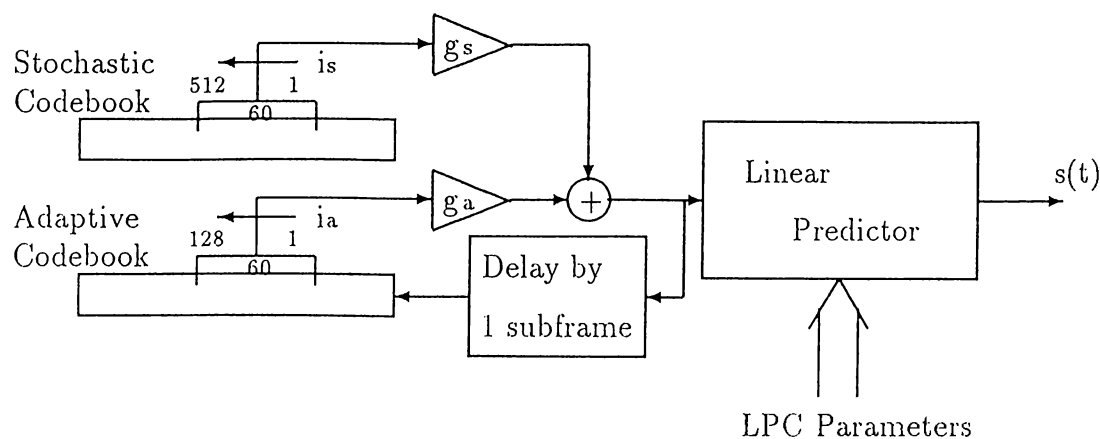
Figure 3.1: *CELP Synthesizer*

Stochastic codebook contains sparse overlapping, ternary valued, pseudo randomly generated codewords. In the stochastic codebook, codewords are overlapped by a shift of -2. In other words each codeword contains all but two samples of the previous codeword and two new samples. The adaptive codebook is a shifting storage register which is updated at the start of each subframe with the previous 60 element LP filter excitation. In the adaptive codebook, codewords are overlapped by a shift of -1. Stochastic codebook vector, which is given by index i_s and scaled by g_s and adaptive codebook vector, which is given by index i_a and scaled by g_a add up to form the linear prediction filter's excitation. Furthermore, the adaptive codebook is updated by this excitation sequence for use in the following subframe.

3.2   Analysis

The CELP analyzer, shown in Figure 3.2, contains a CELP synthesizer and a feedback loop for minimizing the M.S.E. between the original and the synthesized speech. The search procedure finds the adaptive and stochastic indices and gains that minimize the M.S.E.. Codebook search methods for both stochastic and adaptive codebooks are identical. So a two stage search algorithm is used for reducing the complexity of the search. In the first stage the adaptive codebook search which forms the periodic nature of the speech signal is carried out. In the second stage, stochastic codebook search is performed in order to model the random nature of the speech.
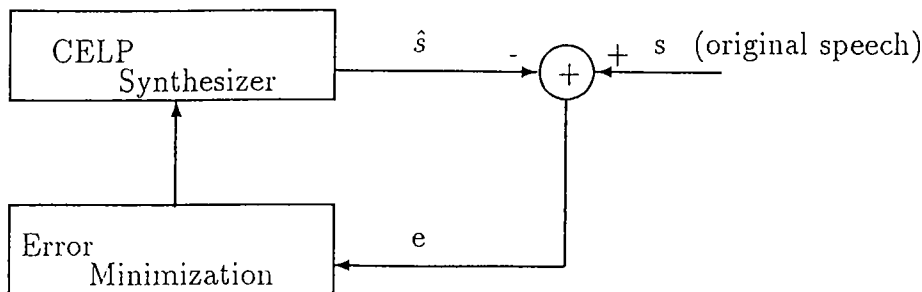
Figure 3.2: *CELP Analyzer*

3.3   Search Algorithm

In this section a brief outline of the codebook search algorithm is presented. Let s, $\hat{s}$, and e represent the original speech, the synthetic speech and the error signal respectively. Let v be the excitation vector being searched for in the present stage and u be the excitation vector of the previous stage. For the first stage the vector u is a zero vector. The excitation vector, v_i, can be written as:

$$v_i = g_i x_i, \tag{3.1}$$

where x_i is the i-th element of the codebook and g_i is the corresponding gain for x_i. Let H be LxL matrix whose j-th row contains the truncated impulse response caused by a unit impulse $\delta(t-j)$ of the LP filter. Here L is the size of the codeword which is also equal to the size of a subframe. Then the synthetic speech can be expressed as the sum of LP filter's zero input response and the convolution of the LP filter's excitation with impulse response.

$$\hat{s}_i = \hat{s}_0 + (u + v_i)H, \qquad i = 1, 2, ..., N \tag{3.2}$$

where u is a zero vector in the first stage search or the scaled adaptive excitation vector in the second stage search. Then the error signal, e_i, is given as follows:

$$e_i = s - \hat{s}_i \tag{3.3}$$

$$= e_0 - v_i H \tag{3.4}$$

where e_0 is the target vector, i.e.,

$$e_0 = s - \hat{s}_i - uH. \tag{3.5}$$

Thus, the error, e_i, can be rewritten as follows,

$$e_i = e_0 - g_i y_i \tag{3.6}$$

where y_i represents the filtered codeword, i.e.,

$$y_i = x_i H. \tag{3.7}$$

Let E_i represents the total square error for codeword i, i.e.,

$$E_i = ||e_i||^2 = e_i e_i^T \tag{3.8}$$

$$= e_0 e_0^T - 2g_i e_0 y_i^T + g_i^2 y_i y_i^T \tag{3.9}$$

The total square error, E_i, is a function of g_i and index i. For optimal gain we set the partial derivative of E_i with respect to g_i to zero, i.e.,

$$\frac{\partial E_i}{\partial g_i} = -2e_0 y_i^T + 2g_i y_i y_i^T = 0. \tag{3.10}$$

Then optimal gain reduces to:

$$g_i = \frac{e_0 y_i^T}{y_i y_i^T} \tag{3.11}$$

after the gain quantized, $\hat{g}_i = Q[g_i]$ the match score can be written as follows,

$$match_i = \hat{g}_i (2e_0 y_i^T - \hat{g}_i y_i y_i^T). \tag{3.12}$$

The search algorithm maximizes this match score for the optimal codeword.

3.4 Implementation of CELP Vocoder on SUN-Sparc Stations

In this section some details of the implementation of 4.8 kbits/sec CELP vocoder is presented. The computational load of the codebook search modules make the real time implementation of the CELP vocoder harder than LPC vocoder. We implemented the 4.8 kbits/sec CELP vocoder on SUN-Sparc Stations. The SUN-Sparc Stations have an input-output (A/D, D/A) channel which uses μ-law compression and stores the speech signal in 8 bits/sample. The *soundtool* software helps in A/D and D/A conversion and supplies speech data for processing. The stored discrete-time speech signal is processed by a software which implements the 4.8 kbits/sec CELP vocoder and synthesized speech is stored in a file. The relative performance of 4.8 kbits/sec CELP vocoder is better than the LPC vocoder. 4.8 kbits/sec CELP vocoder transmits two codebook indices for stochastic and adaptive codebook, and two gain factors for these codebooks, respectively. Also 10 line spectrum pairs are transmitted as the vocal tract parameters. We examine the properties of line spectrum pairs and develop a new coding scheme for these parameters in Chapter 4. The developed software is available at the Electrical and Electronics Engineering Department Library.

Chapter 4

Interframe Differential Coding of Line Spectrum Pairs

This chapter presents a new coding scheme for Line Spectrum Pairs (LSP's). The vocal tract parameters or the LPC coefficients can be represented by the Line Spectrum Pairs which were first introduced by Itakura [12]. For a minimum phase m^{th} order LPC polynomial,

$$A_m(z) = 1 + a_1 z^{-1} + \cdots + a_m z^{-m} \tag{4.1}$$

one can construct two $(m + 1)^{st}$ order polynomials, $P_{m+1}(z)$ and $Q_{m+1}(z)$, by setting the $(m + 1)^{st}$ reflection coefficient to 1 or -1. This is equivalent to setting the corresponding acoustic tube model completely closed or completely open at the $(m + 1)^{st}$ stage. The LSP polynomials, $P_{m+1}(z)$ and $Q_{m+1}(z)$, are defined as follows,

$$P_{m+1}(z) = A_m(z) + z^{-(m+1)} A_m(z^{-1}), \tag{4.2}$$

and

$$Q_{m+1}(z) = A_m(z) - z^{-(m+1)} A_m(z^{-1}). \tag{4.3}$$

It is obvious that $P_{m+1}(z)$ is a symmetric polynomial and $Q_{m+1}(z)$ is an anti-symmetric polynomial. There are three important properties of $P_{m+1}(z)$ and $Q_{m+1}(z)$:

(i) All of the zeros of the LSP polynomials are on the unit circle,

(ii) the zeros of the symmetric and anti-symmetric LSP polynomials are interlaced, and

(iii) the reconstructed LPC all-pole filter maintains its minimum phase property, if the properties (i) and (ii) are preserved during the quantization procedure.

As the roots of $P_{m+1}(z)$ and $Q_{m+1}(z)$ are on the unit circle, the zeros of $P_{m+1}(z)$ and $Q_{m+1}(z)$ can be represented by their angles which are called the LSP frequencies.

In speech compression, the LPC filter coefficients $\{a_i\}_{i=1}^m$ are known to be inappropriate for quantization because of their relatively large dynamic range and possible filter instability problems. The LSP representation of spectral information has both well-behaved dynamic range and filter stability preservation property. Therefore it can be used to encode the LPC spectral information more efficiently than reflection coefficients of the LPC filter [10].

4.1   Computation of LSP Frequencies

In this section, computation of the LSP frequencies is examined, and one method is presented for the real time computation of the LSP frequencies.

As described in the previous section the polynomial $P(z)$ ($Q(z)$) is a symmetric (anti-symmetric) polynomial, i.e.,

$$P(z) = 1 + p_1 z^{-1} + \cdots + p_1 z^{-m} + z^{-(m+1)} \tag{4.4}$$

and

$$Q(z) = 1 + q_1 z^{-1} + \cdots - q_1 z^{-m} - z^{-(m+1)} \tag{4.5}$$

If the order m is even then the polynomials $P(z)$ and $Q(z)$ have the roots $+1$ and -1, respectively which can be removed by a polynomial division. We get two new polynomials,

$$G_1(z) = \frac{P(z)}{1 + z^{-1}} \tag{4.6}$$

and

$$G_2(z) = \frac{Q(z)}{1 - z^{-1}}. \tag{4.7}$$

Let the order of the polynomials $G_1(z)$ and $G_2(z)$ be $2n$, $2n = m$. Let us represent the polynomials $G_1(z)$ and $G_2(z)$ as follows,

$$G_1(z) = 1 + g_1(1)z^{-1} + \cdots + g_1(n)z^{-n} + \cdots + g_1(1)z^{-(2n-1)} + z^{-2n} \tag{4.8}$$

and

$$G_2(z) = 1 + g_2(1)z^{-1} + \cdots + g_2(n)z^{-n} + \cdots + g_2(1)z^{-(2n-1)} + z^{-2n}. \tag{4.9}$$

Polynomials, $G_1(z)$ and $G_2(z)$, contribute n pairs of conjugate zeros and the linear phase term can be removed to give two zero phase series expansion in cosines, i.e.,

$$G_1(e^{-j\omega}) = e^{-j\omega n}G_1'(\omega) \tag{4.10}$$

and

$$G_2(e^{-j\omega}) = e^{-j\omega n}G_2'(\omega) \tag{4.11}$$

where

$$\begin{aligned}G_1'(\omega) &= 2Cos(n\omega) + 2g_1(1)Cos((n-1)\omega) + \cdots \\ &\quad +2g_1(n-1)Cos(\omega) + g_1(n)\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}G_2'(\omega) &= 2Cos(n\omega) + 2g_2(1)Cos((n-1)\omega) + \cdots \\ &\quad +2g_2(n-1)Cos(\omega) + g_2(n)\end{aligned} \tag{4.13}$$

The LSP frequencies are defined as the zeros of $G_1'(\omega)$ and $G_2'(\omega)$. The LSP frequencies (or the zeros) can be found by tracing the frequency ω between 0 and 2π. In real-time implementation the cosine functions bring a heavy computational load. To overcome this problem we can use the frequency mapping $x = Cos\omega$. Then $Cos(m\omega) = T_m(x)$ where $T_m(x)$ is an mth order Chebyshev polynomial in x. Now one can represent (4.12) and (4.13) in terms of Chebyshev polynomials [13] as follows,

$$G_i'(x) = \sum_{k=0}^{n} c_{i,k}T_k(x), \qquad for \quad i = 1,2. \tag{4.14}$$

Using the backward recurrence relationship [14],

$$b_k = 2xb_{k+1} - b_{k+2} + c_{i,k}, \tag{4.15}$$

where $\{b_k\}$ is a sequence with initial conditions $b_N = b_{N+1} = 0$. Then b_0 and b_2 can be calculated with these initial conditions. $G_i'(x)$ can be expressed in terms of b_0 and b_2 as follows,

$$\begin{aligned}G_i'(x) &= \sum_{k=0}^{n}[b_k - 2xb_{k+1} + b_{k+2}]T_k(x) \\ &= \frac{b_0 - b_2 + c_{i,0}}{2}, \qquad i = 1,2.\end{aligned} \tag{4.16}$$

This computation results in a numerically stable evaluation of the Chebyshev polynomial series. The search proceeds backwards from $x = 1$ to $x = -1$. The location of a zero is detected if a sign change occurs in $G_1'(x)$. Once a

zero of the $G_1'(x)$ is found then we search the zero of $G_2'(x)$. This is due to the interlacing property of the zeros of $G_1'(x)$ and $G_2'(x)$ polynomials. The algorithm continues in this way by interchanging the roles of the functions as each zero is found. The increment in this search algorithm should be chosen to be less than 0.0015 for a good precision on the zero locations.

4.2 Differential Coding of LSP Frequencies

In this section a new interframe differential coding scheme is presented for the LSP frequencies [15].

Let $A_{10}^n(z)$ be the LPC filter of the nth speech frame. Corresponding to $A_{10}^n(z)$, 10 LSP frequencies can be uniquely defined. Let us denote the ith LSP frequency of the nth frame as f_i^n, $i = 1, 2, ..., 10$. The key idea of our scheme is to estimate the current LSP frequency, f_i^n, from $(i-1)^{th}$ LSP frequency of the nth frame, f_{i-1}^n, and i^{th} LSP frequency of the $(n-1)^{th}$ frame, f_i^{n-1}, and to quantize the error between f_i^n and the estimate, $\hat{f}_i^n$. In this way, we not only exploit the relation between neighboring LSP frequencies but the relation between the LSP frequencies of the consecutive frames as well. The estimate, $\hat{f}_i^n$, of the LSP frequency, f_i^n, is given by

$$\hat{f}_i^n = \begin{cases} a_i^n \Delta_i + b_i^n f_i^{n-1} & i = 1 \\ a_i^n (f_{i-1}^n + \Delta_i) + b_i^n f_i^{n-1} & i = 2, 3, ..., 10 \end{cases} \tag{4.17}$$

where a_i^n's and b_i^n's are the adaptive predictor coefficients and Δ_i is an offset factor which is the average angular difference between the ith and $(i-1)^{th}$ LSP frequencies. The parameter, Δ_i, is experimentally determined. Predictor coefficients a_i^n's and b_i^n's are adapted by the least mean square (LMS) algorithm,

$$\begin{bmatrix} a_i^n \\ b_i^n \end{bmatrix} = \begin{bmatrix} a_i^{n-1} \\ b_i^{n-1} \end{bmatrix} + \alpha_i^{n-1} \begin{bmatrix} f_{i-1}^{n-1} + \Delta_i \\ f_i^{n-2} \end{bmatrix} d_i^{n-1} \tag{4.18}$$

where

$$d_i^{n-1} = Q[f_i^{n-1} - \hat{f}_i^{n-1}], \tag{4.19}$$

and

$$\alpha_i^{n-1} = \frac{\lambda_i}{(f_{i-1}^{n-1} + \Delta_i)^2 + (f_i^{n-2})^2}, \quad 0 < \lambda_i < 2. \tag{4.20}$$

The parameters, λ_i's, are also experimentally determined.

4.3 Quantizer

The predictor defined in (4.14) is used in an ADPCM structure whose quantizer is designed in the M.M.S.E. sense. A well-known method to design quantizers is the generalized-Lloyd algorithm [16]. However, this algorithm usually converges to locally optimum quantizers. Recently simulated annealing based quantizer design algorithms were developed [17],[18],[19], and it was observed that globally optimal solutions can be reached. In this thesis we use the stochastic relaxation algorithm [18]. We observed that stochastic relaxation algorithm produces better results than the generalized-Lloyd algorithm in the M.S.E. sense.

Stochastic relaxation method utilizes a probabilistic technique for finding globally optimal solutions to complex optimization problems. The main idea is to add an element of zero mean noise to each code vector following the centroid computations in each iteration of the generalized Lloyd algorithm. The noise variance (or the temperature) is then reduced monotonously as the iterations progress.

The design algorithm is as follows:

(a) Code vector initializations:
$$y_1^{(1)}, ..., y_N^{(1)}$$
$$m = 1$$
$$D_0 = \infty$$

(b) Nearest neighbor repartition ($i = 1, ..., M$):
$$j = argmin\{||x_i - y_l|| : 1 \le l \le N\}$$
Let $x_i \in R_j$: j^{th} decision region
$$D_m \leftarrow D_m + ||x_i - y_j^{(m)}||^2$$

(c) Stopping criterion:
If $(D_{m-1} - D_m)/D_m < \epsilon$ stop
else $m = m + 1$

(d) Centroid computation ($i = 1, ..., N$):
$$y_i^{(m)} = \frac{1}{|R_i|} \sum_{x_i \in R_i} x_i$$

(e) Code vector jiggling ($i = 1, ..., N$):
$$y_i^{(m)} \leftarrow y_i^{(m)} + S_i(T_m)$$

goto (b) where S_i is the perturbation value which is a pseudo-random

number generated from a uniform distribution with zero mean and variance, T_m. There are various cooling schedules for T_m, we use the one described in [18],

$$T_m = \frac{\sigma_x^2}{\sqrt{(m+1)}} \qquad (4.21)$$

where σ_x^2 is the input variance.

The M.M.S.E. quantizer, which is designed by stochastic relaxation algorithm, is also scaled during coding by using a backward adaptation structure. Let the current variance of the nth frame and i^{th} LSP frequency be $^n\sigma_i^2$, which is derived from quantizer output as follows

$$^n\sigma_i^2 = \beta_i \, ^{n-1}\sigma_i^2 + (1 - \beta_i)(^{n-1}d_i)^2 \qquad (4.22)$$

where β_i's are experimentally determined constants. In adaptation process each level of the ith quantizer is multiplied by the factor, $\sqrt{\frac{^n\sigma_i^2}{\varrho_i}}$, where ϱ_i is also an experimentally determined constant for the ith quantizer.

4.4   Simulation Examples

In this section we present simulation examples and compare our results to other LSP frequency coding schemes, including the scalar (vector) quantizer based method of Soong and Juang [20] (Farvardin [21], Atal [22]).

The offset factors, Δ_i's were estimated from a training set of 1200 speech frames. The M.M S.E quantizer was trained in a set of 1500 speech frames containing three male and three female persons. The performance of the interframe LSP coding scheme was measured in a set of 3500 speech frames obtained from utterances of three male and three female persons.

We call our LSP coding scheme an interframe method because we not only use the current frame but also the previous frame to code the LSP frequencies of the current frame. A recent method by Soong and Juang which quantize the intraframe differences of the consecutive LSP frequencies f_i^n and f_{i-1}^n reached better results than other scalar quantizers for LSP frequency coding methods [20]. We compare our method to Soong and Juang's method.

Soong and Juang used the log spectral distance distortion measure,

Table 4.1: *Spectral Distortion (SD) Performance of Intraframe and Interframe Coding Schemes*

TOTAL BITS / FRAME	intraframe dB^2 [20]	interframe dB^2
24	3.00	1.72
25	2.60	1.45
26	2.30	1.33
27	2.00	1.16
28	1.80	1.02
29	1.60	0.91
30	1.40	0.87

$d(A(\omega), A'(\omega))$, which is defined in dB^2 as follows

$$d(A(\omega), A'(\omega)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} [B(\omega)]^2 d\omega \qquad (4.23)$$

where $A(\omega)$ and $A'(\omega)$ are the original and the reconstructed LPC frequency responses respectively, and $B(\omega)$ is given by,

$$B(\omega) = 10log\frac{1}{|A(\omega)|^2} - 10log\frac{1}{|A'(\omega)|^2} \qquad (4.24)$$

$B(\omega)$ is called the log spectral difference.

In Table 4.1 average log spectral distances for total number of bits used to code a set of LSP frequencies, f_i, $i = 1, 2, ..., 10$. are given. In the second column of Table 1 coding results given in [20] are summarized. In third column, coding results of our method are described.

We also compare our method to vector quantizer based methods. Farvardin [21], and Atal [22] reached 1.0 dB spectral distortion at 24 bit/frame rate with vector quantizers. By using Huffman coding of the quantizer output, we reached lower bit rates for 1.0 dB spectral distortion in our method.

In these comparisons we consider the following distortion measure

$$d'(A(\omega), A'(\omega)) = [\frac{1}{2\pi} \int_{-\pi}^{\pi} [B(\omega)]^2 d\omega]^{1/2} \qquad (4.25)$$

which is used in [21], [22].

We used 6, 7, 8, and 9 level quantizers for Huffman coding, and obtained a single Huffman codebook for all 10 LSP frequencies. Table 2-3, give the results of Farvardin [21], Atal [22], and our method with Huffman coding, respectively. In terms of spectral distortion, our method is better than [21] and [22]. However percentage outliers of VQ based methods is lower than our method which is also well within the acceptable range, i.e., our method has < 2% outlier frames in the range 2-4 dB and has no outlier with spectral distortion > 4 dB [22].

Table 4.2: *Spectral Distortion (SD) Performance of the Vector Quantizers [21] and [22]*

Rate bits/frame	Farvardin [21]		Atal [22]	
	Average SD (dB)	Outliers > 2dB(%)	Average SD (dB)	Outliers > 2dB(%)
22	-	-	1.17	2.73%
23	-	-	1.10	1.60%
24	1.11	1.50%	1.03	1.03%
25	1.02	0.20%	0.96	0.61%
26	0.97	0.05%	0.90	0.44%
27	0.94	0.02%	-	-

Table 4.3: *Spectral Distortion (SD) Performance of the Interframe Differential Coding with Entropy Coding*

Quantizer level	Average Rate(bits/frame)	Average SD (dB)	Outliers > 2dB(%)
6	23.45	1.04	2.80%
7	24.25	0.91	1.50%
8	26.55	0.82	0.77%
9	27.66	0.76	0.61%

Although we used different evaluation data sets than [20], [21] and [22] (The sets used in [20], [21] and [22] are also different from each other), we conclude the following points from our simulation examples. We observe that interframe differential coding of LSP frequencies is more advantageous than scalar intraframe coding. This improvement is achieved by slightly increasing the computational complexity of the coder. The performance of our coding

method is comparable to vector quantizer based methods and the computational complexity of our coder is much lower than vector quantizer based methods.

Chapter 5

Conclusion

In this thesis, low bit rate speech coding techniques are examined. LPC vocoder is the earliest low bit rate speech coding method [1]. But the performance of the LPC vocoder depends on several important parameters such as, vocal tract parameters, pitch period, and gain and a good combination and continuity of these parameters. Simulations of the LPC vocoder showed that smoothing of these parameters in consecutive frames is necessary as well as good estimates of these parameters. The most important parameter is the pitch period. It is difficult to establish an error measure for pitch period estimation. However we observed that cepstrum method achieves better performance than AMDF method in simulations.

A recently developed low bit rate speech coder is the CELP vocoder which also uses linear prediction, but the main difference of this method is the analysis-by-synthesis search procedures which form a closed-loop system. Stochastic and adaptive codebook sizes are the most effective parameters for the performance of the CELP vocoder. Some implementations use 128 integer and 128 non-integer delays for adaptive codebook and this increase the performance of vocoder especially for female speech signal. But enlarging the codebook size creates a trade off with the increase in computational load of the search procedures.

In this thesis we also developed a new coding scheme for vocal tract parameters. The new interframe differential coding scheme outperformed the scalar quantizer based methods and reached the performance of vector quantizer based methods. This improvement is achieved by slightly increasing the complexity of the scalar coders. The new LSP frequency coding method can be used both in LPC and CELP vocoders. This will result lower bit rates than ordinary LSP frequency coding methods.

28

An interframe vector quantization based differential coding scheme for LSP frequencies can be developed as a future work. This will increase the computational complexity of our scheme however we believe that the performance will be better than other vector quantizer based methods.

References

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

[2] V.Cuperman B.S. Atal and A. Gersho. *Advances in Speech Coding.* Kluwer Academic Publishers, 1991.

[3] Halil Padır. *An LPC Vocoder System.* M.Sc. thesis, METU, 1983.

[4] B.S. Atal and S.L. Hanauer "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Am.*, pp. 637–655, 1971.

[5] J. Makhoul "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, 1975.

[6] C.K. Un and S.C. Yang "A pitch extraction algorithm based on lpc inverse filtering and amdf," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 25, pp. 565–572, December 1977.

[7] A. Cohen, R. Freudberg, M.J. Ross, H.L. Shaffer and H.J. Manley "Average magnitude difference function pitch extractor," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 22, pp. 353–362, October 1974.

[8] A.E. Rosenberg, L.R. Rabiner, M.J. Cheng and C.A.McGonegal "A comparative performance study of several pitch detection algorithms," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 24, pp. 399–418, October 1976.

[9] Military Agency for Standardization. "NATO stardardization agreement, Stanag 4196, parameters and coding characteristics that must be common to assure interoperability of 2400 bps linear predictive encoded digital speech,".

[10] National Communications System Office of Technology and DC Standards, Washington. "Proposed federal standard 1016, analog to digital conversion

of radio voice by 4800 bit/second code excited linear prediction (CELP),",
September 1989.

[11] V.C. Welch, J.P. Campbell and T.E. Tremain. "The new 4800 bps voice
coding standard,". Military and Government Speech Tech'89 4800 bps
Voice Coding Session, Arlington, Virginia, November 1989.

[12] F. Itakura "Line spectrum representation of linear predictive coefficients
of speech signals," *Journal of Acoust. Soc. Am.*, p. 535a, 1975.

[13] P. Kabal and R.P. Ramachandran "The computation of line spectral
frequencies using chebyshev polynomials," *IEEE Trans. on Acoustics,
Speech, and Signal Processing*, vol. 34, pp. 1419–1426, December 1986.

[14] Kendall E. Atkinson. *An Introduction to Numerical Analysis*, chapter 4,
pp. 221–222. Wiley, 1988.

[15] E. Erzin and A. E. Çetin "Interframe differential coding of line spec-
trum pairs," presented in *26-th Conference on Information Sciences and
Systems, Princeton*, March 1992.

[16] A. Buzo Y. Linde and R.M. Gray "An algorithm for vector quantizer
design," *IEEE Trans. on Comunications.*, vol. 28, pp. 84–95, January
1980.

[17] A. Enis Çetin and V. Weerackody "Design of vector quantizers using
simulated annealing," *IEEE Trans. on Circuits and Systems*, vol. 35, p.
1550, 1988.

[18] K. Zeger and A. Gersho "Stochastic relaxation algorithm for improved
vector quantiser design," *Electronics Letters*, vol. 25, pp. 896–898, July
1989.

[19] J. Vaisey, K. Zeger and A. Gersho "Globally optimal vector quantizer de-
sign by stochastic relaxation," *IEEE Trans. on Signal Processing*, vol. 40,
pp. 310–322, February 1992.

[20] F. Soong and B.H. Juang. "Optimal quantization of lsp parameters,".
accepted for publication in *IEEE Trans. on Signal Processing*.

[21] N. Phamdo, R. Laroia and N. Farvardin "Robust and efficient quantization
of lsp parameters using structured vector quantizers," *Proc. of the Int.
Conf. on Acoustics, Speech and Signal Processing 1991 (ICASSP '91)*, pp.
641–645, May 1991.

[22] K.K. Paliwal and B.S.Atal "Efficient vector quantization of lpc parameters at 24 bits/frame," *Proc. of the Int. Conf. on Acoustics, Speech and Signal Processing 1991 (ICASSP '91)*, pp. 661–664, May 1991.

[23] S. Singhal, Private communications, 1991.