Switched Lattice-Based Quantization of LSF Parameters

Ahmad Fazel and A. K. Khandani

Department of Electrical and Computer Engineering

University of Waterloo

Waterloo, Ontario, Canada, N2L 3G1

Technical Report UW-E&CE#99-14

November 22, 1999

Abstract

In this work, a switched lattice-based quantizer for the quantization of speech LSF parameters is presented. LSF difference values are first quantized using a scalar quantizer, as well as a two dimensional vector quantizer, and subsequently adjusted for a lower bit-rate using a lattice. A Euclidean distance measure is used as the distortion measure for the design, while a spectral distortion measure is used for the evaluation of the performance. LSF differences are formed in a closed loop manner for which a first-order prediction filter is used. The parameters for the prediction filter are calculated from a LSF database. Lattice-based double frame and single frame quantization is performed for each frame and the one which results in a lower distortion is chosen. Numerical results are presented showing an excellent performance with very low complexity. The results are compared to the split vector quantizer using common database for test and training showing substantial improvement. Finally, the issue of the robustness to channel errors is investigated.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Speech coding is concerned with the compact representation of digital voice signals. The objective in speech coding is to represent the speech signal with a minimum number of bits while maintaining its perceptual quality. The representation can be direct or *parametric.* Speech is typically band limited to 4 kHz and sampled at 8 kHz. The simplest non-parametric coding scheme is simply a quantization of sampled amplitudes (i.e. Pulse Code Modulation). A typical quantization scheme uses 8 bits per sample, resulting in an overall rate of 64 kbits/sec. Speech coded at 64 kbits/sec is considered to be non-compressed and is often used as a reference for compression. *Medium rate* coding refers to the range of $8-16$ kbits/sec, *low rate* coding to the range below 8 kbits/sec down to 2.4 kbits/sec, and *very low rate* to rates below 2.4 kbits/sec [16]. Speech coding at medium rates and below is achieved through parametric representation. Parametric representation consists of *analysis* and *synthesis*. In the analysis stage, important parameters are extracted from the speech. In the synthesis stage, extracted parameters are used to reconstruct the original speech. Most speech coding schemes at low bit-rates (less than 8 kbits/sec) make use of the Linear Predictive Coding (LPC) model for the speech. Usually, the LPC coefficients are not encoded directly. Instead, equivalent parameters such as reflection coefficients or line spectrum pair (LSP) parameters are quantized for transmission. LSP frequencies were introduced by Itakura in 1975 [3] and are mainly used for the representation of LPC coefficients. In this work, the LSF parameters, which are scaled versions of LSP parameters, are used to represent the LPC coefficients.

1.1.1 CELP Coding

Code Excited Linear Prediction (CELP) was first introduced by Atal and Schroeder [17] as a powerful technique for low bit-rate speech coding. The bit-rate efficiency of the CELP coders is attributed to block-by-block speech processing in contrast to the sample-by-sample approach taken by PCM coders. In this technique, the speech is segmented into blocks of typically 20 ms called a *speech frame*. Based on classical speech production theory [14], in each such frame, speech can be modeled as the output of two cascaded prediction filters, i.e. short-term and long-term predictor filters, driven by an excitation signal. The excitation signal is a representation of the stochastic nature of the glottal excitation and is modeled by a white noise signal. The long-term prediction filter represents the *pitch* of the voiced sounds due to the periodic nature of the vocal cord vibration. The short-term prediction filter, referred to as the linear prediction filter, represents the spectral envelope of the speech and represents the shaping effect of the vocal tract. LPC analysis is performed on a frame-by-frame basis. Afterwards, the best excitation codeword is obtained using an *analysis-by-synthesis* method. The quantized parameters, corresponding to the filter coefficients, and the excitation codebook indices are then encoded for transmission.

To provide a detailed account of the above, the Vector Sum Excited Linear Prediction (VSELP) speech coder which is used in the U.S. digital cellular system (IS-54) is explained in the next section as a practical example.

1.1.2 Vector Sum Excited Linear Prediction (VSELP)

The Vector Sum Excited Linear Prediction (VSELP) speech coder employs a structure which allows for a very efficient search procedure of its codebooks. Figure (1.1) shows the block diagram of VSELP, where p_n is the original speech signal. This process involves stepping through the codebook entries to find the code vector which generates the least difference error signal, e_n.

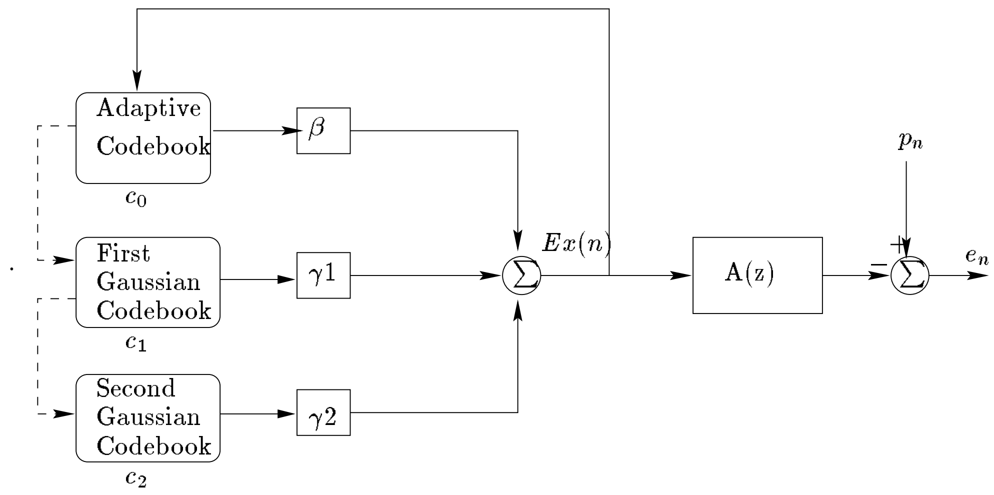

Figure 1.1: Block Diagram of VSELP

This coder operates at a bit rate of 7950 bits/sec. In VSELP, speech is segmented into 20 ms frames. This corresponds to 160 samples of speech and each frame is further divided into four sub-frames. In this case,38 bits are used for the representation of LPC coefficients of order 10. The bit allocations for these coefficients are nonuniform. The number of allocated bits for $LPC_1, LPC_2, \ldots, LPC_{10}$ are $6, 5, 5, 4, 4, 3, 3, 3, 3, 2$ bits, respectively. Three codebooks for the excitation vectors are employed. One of them is adaptive and the other two are Gaussian codebooks. There are 128 code vectors in each codebook. Thus, $3 \log_2(128) = 21$ bits are required to transmit the indices of the codebooks in each sub-frame. A gain-shape vector quantizer, which normalizes the energy level of a vector prior to encoding, is employed. Therefore, 5 bits are transmitted per frame for the average speech energy and

8 bits per sub-frame for gain factors. Using these numbers, it is observed that the bit-rate is 7950 bits/sec and 37% of the transmitted information is for the encoding of the LPC coefficients.

1.2 Scalar and Vector Quantization of LSF Parameters

The challenge in the quantization of LSF[1] parameters is to achieve transparent quality with the minimum bit-rate while maintaining the memory and computational complexity as low as possible. Various scalar and vector quantization techniques have been proposed in the literature for quantization of LSF parameters. Each of these techniques offers a trade-off between bit rate and complexity. Scalar quantizers are attractive for their low level of complexity. However, they achieve transparent quality at high rates of 30 to 40 bits per frame [7]. Differential scalar quantization of LSF parameters, which is the quantization of the differences between successive LSFs, has been proposed by Soong and Juang to exploit the correlation between successive LSFs [13]. By differential quantization, the correlation between the LSF parameters within each frame is exploited. This correlation is called *intra-frame* correlation. The results show that a saving of about 2 bits per

[1] Refer to chapter 2 for the definition of LSF parameters and their relation with LPC coefficients.

frame can be achieved using differential quantization. However, a drawback is that differential quantization is more vulnerable to channel noise.

The optimal solution to spectral coding is a vector quantizer which exploits the joint statistical properties of the LSF parameters. However, the high complexity of this method makes it impractical. A full search vector quantizer is estimated to achieve transparent quality at 18 bits per frame; however it requires a codebook with over one million floating point numbers. Sub-optimal vector quantization techniques such as multi-stage vector quantization, hybrid scalar-vector quantization and split vector quantization have been proposed in the literature to reduce the complexity associated with vector quantizers at the cost of a degradation in performance. A split vector quantization approach which was originally introduced by Paliwal and Atal [12][2] is simulated and used as the basis for the comparison in this work.

Besides the intra-frame correlation of the LSF parameters, *inter-frame* correlation of LSF parameters can be exploited to reduce the corresponding bit-rate. This means that the LSF parameters of a given frame can be predicted from the LSF parameters of previous frames. Since all inter-frame encoders use the information of the previous frames, they suffer from error propagation in communication over noisy channels. In this work, a double frame quantizer is employed which exploits the correlation of the two adjacent frames using a first-order linear predictor.

[2]This is the common method of quantization in most of the current standards.

1.3 Goal of this Work

In this work, a new technique for the quantization of LSF parameters is presented. As shown in the VSELP coder, about 37% of the transmitted information is dedicated to the LPC coefficients (LSF parameters). Therefore, it is very important to represent the LSF parameters with as few bits as possible while maintaining the highest speech quality with minimum memory requirements and computational complexity.

It is the goal of this work to achieve comparable performance to that of a split vector quantizer[12] with very low memory requirements and computational complexity.

1.4 Prefiltering

The original speech signal should be passed through a high-pass filter to obtain a similar dynamic range over the entire frequency range. Passing the original signal through a high-pass filter should be done before extracting the LPC parameters. Since different kinds of high-pass filters can be employed, different sets of LPC parameters can be produced.

The performance of a LSF quantizer depends on the training database it uses. Therefore, it is not fair to compare the performance of different quantizers when they do not use a common training and test database. Because of this fact, a limited comparison is done in this work. For the basis of comparison, a split vector LSF quantizer is designed. The database that is used contains approximately $175,000$ LSF vectors to train the quantizer and $11,000$ LSF different test vectors to test the performance.

1.5 Organization of the Report

In the next chapter, the principles of Linear Predictive Coding and Line Spectral Frequencies are presented. A spectral distortion measure is used as an objective measure to assess the quality of LSF quantizers is also discussed. In Chapter 3, scalar lattice quantization and two dimensional vector lattice quantization are introduced. Trellis-based decoding and the Viterbi algorithm are presented as a computationally efficient search technique. Euclidean distance and weighted Euclidean distance measures which are used as a replacement for the spectral distortion measure to reduce the complexity are discussed in this chapter as well. In Chapter 4, the main structure of the proposed quantizer, the Switched Lattice-Based Quantizer and also the differential encoding technique employed in this work are explained. The complexity and performance results are presented in Chapter 5 the along with the comparisons with the spilt vector quantizer. It is shown that by using the proposed switch lattice-based quantizer a comparable performance is achievable with much lower complexity and memory requirements.

Chapter 2

Linear Prediction and LSF Parameters

2.1 LPC Analysis

A frame of a speech signal consists of N samples, say $s_1, s_2, \ldots, s_N$, where N is equal to 160 for the speech signal sampled at the rate of 8 kHz and a frame size of 20 ms. Each sample can be predicted from the p previous samples using,

$$\tilde{s}(n) = \sum_{i=1}^{p} a_i \; s(n-i) \tag{2.1}$$

where p is called the order of prediction and is typically equal to 10. The coefficients $a_1, a_2, \ldots, a_p$ are called the LPC coefficients. The prediction error

$e(n) = s(n) - \tilde{s}(n)$ is equal to,

$$e(n) = s(n) - \sum_{i=1}^{p} a_i \; s(n-i) \tag{2.2}$$

and the average prediction error energy is

$$E[e^2(n)] = E[(s(n) - \sum_{i=1}^{p} a_i \; s(n-i))^2]. \tag{2.3}$$

To calculate the LPC coefficients, the partial derivatives of the above equation with respect to the coefficients a_i are set to zero,i.e.,

$$\frac{\partial E[e^2(n)]}{\partial a_i} = 0. \tag{2.4}$$

In this way, we obtain p equations and solving them results in the values of the LPC coefficients.

Assuming that the speech signal is stationary in a frame interval, we have

$$E[s(n-i)s(n-j)] = r(|i-j|) \tag{2.5}$$

where $r(k)$ is the autocorrelation of the sequence $s(n)$, and is defined as:

$$r(k) = \sum_{n=0}^{N} s(n)s(n-k). \tag{2.6}$$

The above p equations, which are called the Yule-Walker equations, can be written in the matrix form:

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r_(p) \end{bmatrix} \tag{2.7}$$

The matrix

$$R = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix} \tag{2.8}$$

is called the autocorrelation matrix and not only has symmetry but also has a *Toeplitz* structure. The matrix R, due to its structure, can be inverted by a computationally fast algorithm such as the Levinson-Durbin algorithm to compute the LPC coefficients.

The linear predictive coding (LPC) analysis results in an all-pole filter $H(z) = 1/A(z)$, where

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_p z^{-p}. \tag{2.9}$$

The Toeplitz structure also guarantees that the poles of the filter $H(z)$ are inside the unit circle, meaning that the synthesis filter $H(z)$ is stable [5].

2.2 Line Spectral Pair Parameters

In speech-coding applications, it is required that the all-pole filter remains stable after quantization of the LPC parameters. Since quantization error in direct scalar quantization of the LPC parameters produces relatively large spectral distortion, instability of the filter $H(z)$ can result. It is also computationally expensive to ensure the stability of the filter $H(z)$ using LPC parameters. Because of these problems, it is necessary to transform the LPC coefficients to another representation which is more suitable for quantization. In the literature, a number of such representations have been proposed. Among them, line spectral frequency (LSF) is the most widely used representation of LPC coefficients. The LSF representation has a number of desirable properties, including a bounded range, a sequential ordering of the parameters and a simple check for the filter stability.

The filter $A(z)$ defined by (2.9) may be decomposed to:

$$P(z) = A(z) + z^{-(p+1)}A(z^{-1})$$

$$Q(z) = A(z) - z^{-(p+1)}A(z^{-1}) \tag{2.10}$$

The LSF polynomials, i.e. $P(z)$ and $Q(z)$, are of order $p+1$ and thus have $p+1$ zeroes each. From their total of $2p+2$ zeros, two of them are located at $z=1$ and $z=-1$.

The LPC analysis filter is reconstructed using,

$$A(z) = \frac{1}{2}[P(z) + Q(z)]. \tag{2.11}$$

All roots of $P(z)$ and $Q(z)$ lie on the unit circle, figure 2.1, and they alternate with each other (ordering property).

Since zeros of $P(z)$ and $Q(z)$ are on the unit circle, they can be represented as $e^{j\omega}$ and the ω's are then called Line Spectral Pair frequencies (LSPs). The Line Spectral Frequencies are defined as a scaled version of the LSPs defined

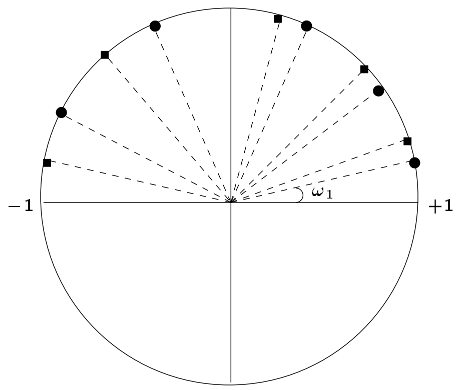

Figure 2.1: 10^{th} order LSF parameters representation on the unit circle

by:

$$lsf_i = \frac{1}{2\,\pi}\, lsp_i, \quad i = 1, \ldots, p. \tag{2.12}$$

Since the polynomial coefficients are real, the roots occur in complex conjugate pairs, and the angles appear in pairs. Thus, the angles between $0 < \omega_i < \pi$ are sufficient to represent the original LPC parameters. Therefore, the LSFs are ordered frequencies between 0 and 0.5. In this work, a 10th order LPC analysis for each frame, i.e. 20 ms of speech, is employed.

It has been proved that the ordering property is a necessary and sufficient condition for the stability of the LSF synthesis filter [3]. Therefore, to guarantee stability for the LSF synthesis filter, we must make sure that the quantized versions of LSF parameters also satisfy the ordering property. The other important property of LSF parameters is localized spectral sensitivity. Therefore, quantization error on one of the LSF values just affects its own spectral neighborhood.

2.3 Spectral Distortion

Ideally, the quality of an LSF quantizer should be judged by means of subjective listening using human test subjects. In practice, however, an objective measure is used for the LSF quantization procedure. Spectral distortion as defined in [12] is normally used to assess the performance of the quantizers.

Assume that the un-quantized and quantized LSF parameters result in the reconstruction filters $A(z)$ and $\hat{A}(z)$, respectively. Then, the power spectral densities are given by,

$$\begin{aligned} P(f) &= \frac{1}{|A(\exp(j2\pi f/f_s)|^2} \\ \hat{P}(f) &= \frac{1}{|\hat{A}(\exp(j2\pi f/f_s)|^2} \end{aligned} \tag{2.13}$$

where f_s is the sampling frequency.

The Spectral Distortion (SD) of a given frame is defined as,

$$D = \sqrt{\frac{1}{f_2 - f_1} \int_{f_1}^{f_2} [B(f)]^2 df} \tag{2.14}$$

where

$$B(f) = 10\log_{10}[P(f)] - 10\log_{10}[\hat{P}(f)].$$

The spectral distortion is evaluated for all frames in the test data and its average value represents the distortion associated with a particular quantizer.

If LSF quantization does not introduce any additional audible distortion the quantization is *transparent*. For spectral transparency an average distortion of less than 1dB is required. However, it has been observed [16] that too many *outlier* frames having large spectral distortion can cause audible distortion, even though the average spectral distortion is less than 1dB. A frame is defined to be an outlier frame if it has a spectral distortion greater than 2dB. Transparent quality is considered as the average spectral distortion of about 1dB and less than 2% outliers above 2 dB [12].

Chapter 3

Lattice-Based Quantization

3.1 Lattice Quantization

A lattice is a discrete set of N dimensional vectors which forms a group. The fundamental region of a lattice is a building block which fills the whole space with just one lattice point in each copy. Lattice quantizers are based on using the points of the lattice to partition the space into the quantizer regions. We can achieve a reduction in the quantization noise by using the points of lattice instead of an N dimensional rectangular array of points (quantization gain). For decoding, given a point $\mathbf{r} = (r_1, \ldots, r_N)$, we find a point $\mathbf{x}$ in the lattice which has the minimum Euclidean distance to $\mathbf{r}$. The objective is to choose a lattice with a low decoding complexity and with a high quantization gain. In this work, lattice D_{10} is employed which comprises the points $(x_1, x_2, \ldots, x_{10})$, for which

$$\sum_{i=1}^{10} x_i \equiv 0 \pmod 2. \tag{3.1}$$

If an LBG quantizer with 2^k points is implemented along each dimension (total rate of Nk using an integer grid), then the use of the lattice D_{10} results in a rate of $Nk - 1$ bits.

In this case, using a latticed quantizer with four points along each dimension results in a quantizer of 19 bits and using a vector lattice quantizer of dimension two with 16 reconstruction points results in a quantizer of 39 bits (per two samples).

3.2 Trellis-based Decoding

The structure of a lattice can be represented by a state diagram called a trellis which consists of a number of stages, and each stage is composed of a set of states. There are a number of transitions from any state of a given stage to certain states in the next stage. The transitions are unidirectional and are referred to as branches. The paths in the trellis correspond to quantizer points. Decoding is simplified by finding the path with the least additive distortion through the trellis. The Viterbi algorithm to be considered shortly indeed finds a minimum distortion path in the trellis in an efficient manner. The trellis diagram as shown in Figure 3.1 is employed for this work where k_i represents the index of quantization levels of LSF differences and the

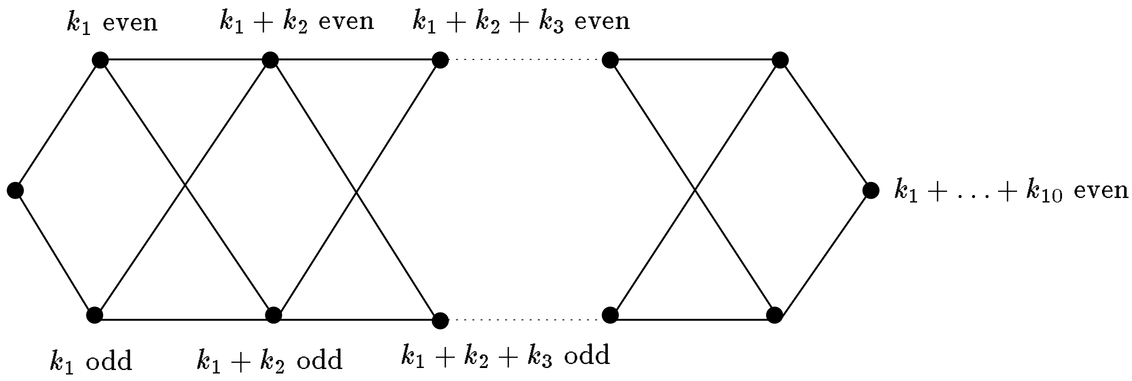

Figure 3.1: Trellis-based decoding of the lattice D_{10}

states correspond to the LSF values. The very first state represents the first component of the LSF vector. It can be quantized using either the quantization levels whose indices are even or quantization levels whose indices are odd. This process goes on until all the LSF vector components are quantized.

3.3 Search Through the Trellis

We must find a path through the trellis which results in a minimal quantization error. The Viterbi algorithm is a minimum cost search technique for a trellis. The basic idea is that, the minimum distortion path from the stage 0 to the stage n must be an extension of one of the minimum distortion paths to a state at stage $n-1$. Thus, in order to find the best possible path, or *survivor path*, we find the best path to each state for each stage by finding the best extension from the previous states into the current state and we perform this for each stage up until the last stage. Figure 3.2 depicts this procedure.

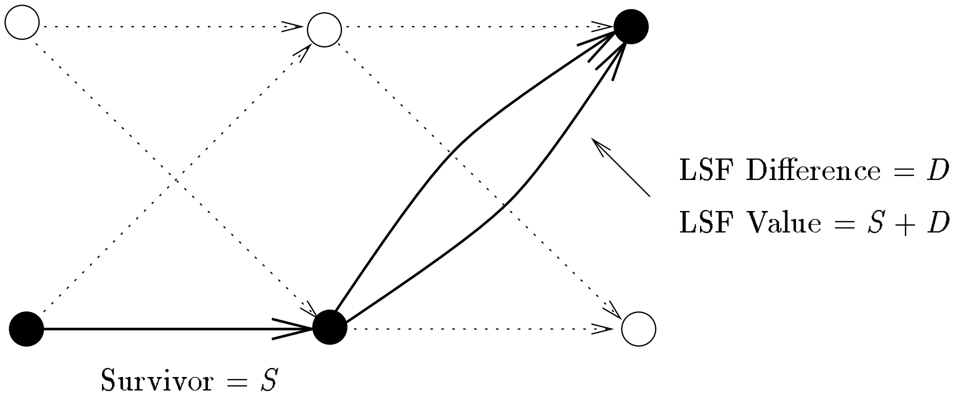

Figure 3.2: Viterbi Algorithm

3.4 Distance Measure

Selection of proper distortion measure is a very important issue in *vector* quantization of LSF parameters. Ideally, the spectral distortion defined by equation (2.14) should be used to design the vector quantizer. Due to the enormous computational complexity associated with the spectral distortion measure, simpler distance measures are employed to design the vector quantizer. A Euclidean distance measure is commonly used as the distance measure between the original and quantized version of LSF parameter vectors. Since different components of an LSF parameter vector have different spectral sensitivities, a weighted Euclidean distance measure which assigns weights to individual LSFs according to their spectral sensitivity is usually employed. The weighted Euclidean distance measure $d(\mathbf{L}, \hat{\mathbf{L}})$ between the LSF parameter vector $\mathbf{L}$ and its quantized version $\hat{\mathbf{L}}$ is given by:

$$d(\mathbf{L}, \hat{\mathbf{L}}) = \sum_{i=1}^{10} c_i w_i (L_i - \hat{L}_i)^2 \tag{3.2}$$

where L_i and $\hat{L}_i$ are the ith LSFs in the original and quantized vector, respectively. The weights c_i do not change frame-to-frame. These weights used to emphasize the lower frequency components which are more important than the higher frequency components. The weights w_i are variable frame-to-frame and are derived from the LSF parameter vector in each frame. We employed constant weight $c_i = 1$ and a nonlinear weight function for variable weights given by:

$$w_1 = \begin{cases} 1.0, & \text{when } \; [2\pi(L_2 - 0.02) - 1] > 0 \\ 10\ [2\pi(L_2 - 0.02) - 1]^2 + 1, & \text{otherwise,} \end{cases}$$

for $1 < i < 10$

$$w_i = \begin{cases} 1.0, & \text{when } \; [2\pi(L_{i+1} - L_{i-1}) - 1] > 0 \\ 10\ [2\pi(L_{i+1} - L_{i-1}) - 1]^2 + 1, & \text{otherwise,} \end{cases}$$

and

$$w_{10} = \begin{cases} 1.0, & \text{when } \; [2\pi(0.471 - L_{10}) - 1] > 0 \\ 10\ [2\pi(0.471 - L_{10}) - 1]^2 + 1, & \text{otherwise.} \end{cases} \tag{3.3}$$

This weight function is based on the idea of emphasizing LSF vector components which are closely located. The values 0.02 and 0.471 are respectively the minimum and maximum value of LSF vectors. This weight function is used in the ITU-T G.729 standard and was recommended by Nortel Networks for this work.

Since the weighted Euclidean distance measure had a negligible performance improvement for lattice quantizers, the Euclidean measure distance is simply used for the design of the single and double frame quantizers. However, in the design of the split vector quantizer, used as a basis for comparison, the weighted Euclidean distance is still employed.

Chapter 4

Switched Single and Double Frame Quantization

The LSF values within a speech frame (LSF vector) are correlated which is referred to as *intra-frame* correlation of the LSF parameters. The LSF vectors from adjacent speech frames are also correlated and is referred to as the *inter-frame* correlation. We can use this fact to predict each LSF value from its past and transmit the difference between the predicted value and its actual value.

In a quantizer, the quantization interval depends on the variance of the input and the quantization intervals determine the quantization noise. By differential encoding, we reduce the variance of the input sequence which in turn decreases the quantization noise.

Let us consider a sequence L_n. A difference sequence d_n is formed as $d_n = L_n - L_{n-1}$. This difference sequence is quantized to obtain the sequence $\hat{d}_n = Q[d_n] = d_n + q_n$, where q_n is quantization error. At the receiver side, the reconstructed sequence $\hat{L}_n$ is obtained by adding $\hat{d}_n$ to the previous reconstructed value $\hat{L}_{n-1}$. The nth element in the sequence at the receiver is:

$$\hat{L}_n = L_n + \sum_{i=1}^{n} q_i \tag{4.1}$$

Notice that, what the encoder quantizes is different from what the decoder sees. The only information available to the receiver is the reconstruction value $\hat{L}_n$. If instead of $d_n = L_n - L_{n-1}$, the transmitter uses $d_n = L_n - \hat{L}_{n-1}$, the receiver and transmitter work on the same information for each element in the sequence:

$$\begin{aligned} \hat{e}_i &= e_i + q_i \\ \hat{L}_{i+1} &= (\alpha_i \hat{L}_i + \beta_i) + \hat{e}_i \\ &= (\alpha_i \hat{L}_i + \beta_i) + e_i + q_i \\ &= L_{i+1} + q_i. \end{aligned} \tag{4.2}$$

For this case, $\hat{L}_n = L_n + q_n$ and there is no accumulation of quantization noise.

4.1 First-order Linear Prediction

Using a first-order predictor, each component of LSF vector can be predicted from the previous component:

$$\hat{L}_{i+1} = \alpha_i L_i + \beta_i. \tag{4.3}$$

The prediction error, $e(n+1) = L(n+1) - \tilde{L}(n+1)$, is

$$e(i+1) = L(i+1) - (\alpha_i L_i + \beta_i) \tag{4.4}$$

and the average prediction error energy is

$$E[e^2(i)] = E[(L(i+1) - (\alpha_i L_i + \beta_i))^2]. \tag{4.5}$$

In order to calculate the coefficient α_i, the partial derivatives of the above equation with respect to the coefficient α_i is set to zero

$$\frac{\partial E[e^2(i)]}{\partial \alpha_i} = 0. \tag{4.6}$$

Solving the above equations, α_i is calculated as,

$$\alpha_i = \frac{COV(L_i, L_{i+1})}{VAR(L_i)}, \tag{4.7}$$

where E is expectation and VAR and COV are variance and covariance, respectively.

From equation (4.3) β_i can be calculated by:

$$\beta_i = E[L_{i+1}] - \alpha_i E[L_i]. \tag{4.8}$$

The calculated α_i and β_i minimize the average prediction error energy, i.e. average mean squared error.

4.2 Differential Encoding

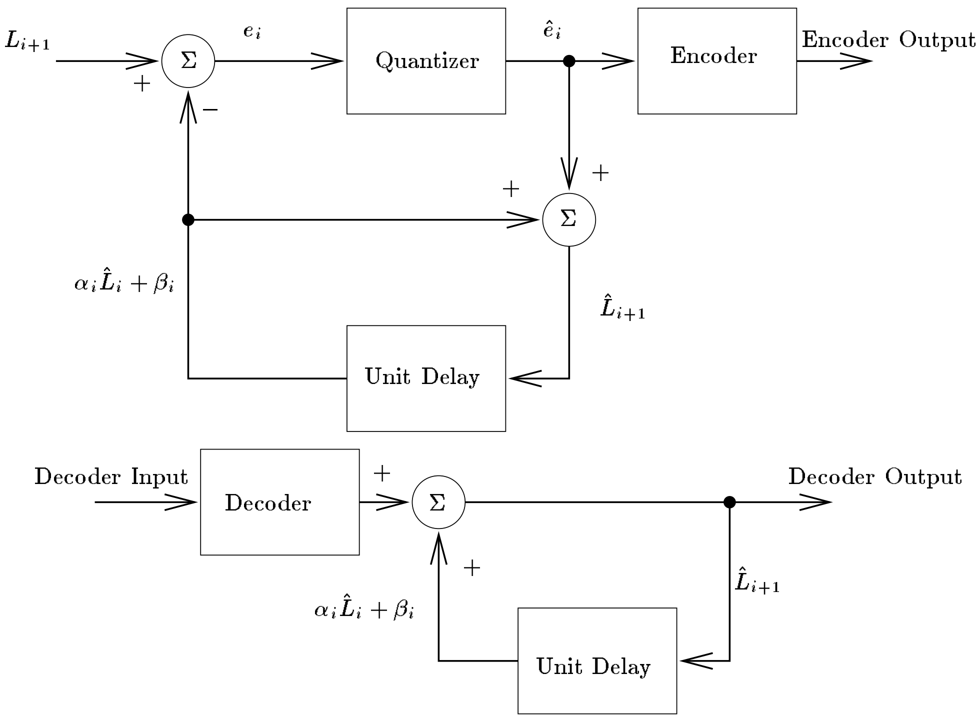

Figure 4.1: Block Diagram of Differential Encoding

The differential encoding procedure employed in this work is summarized below:

(1) Begin by quantizing l_1 to $\hat{l}_1$, $i = 0$;

(2) Form the difference between l_{i+1} and $\alpha\hat{l}_i + \beta_i$, i.e. $\Delta l_i = l_{i+1} - (\alpha\hat{l}_i + \beta_i)$;

(3) Quantize Δl_i to $\Delta\hat{l}_i$;

(4) Reconstruct $\hat{l}_{i+1} = \hat{l}_i + \Delta\hat{l}_i$;

(5) $i \leftarrow i + 1$, if $i > 10$ (LSF order), stop ; Otherwise go to (2).

The database used for the training of the quantizers, is used to calculate the constants α_i and β_i. The corresponding block diagram is shown in the Figure 4.1.

4.3 Single Frame Scalar Quantization

Individual LSF parameters are quantized using a scalar quantizer. The quantizer is designed from the training data set of the Nortel Networks database using the Linde-Buzo-Gray (LBG) algorithm. To exploit the intra-frame correlation of a LSF vector, the differential quantization described in the previous section is employed.

Each LSF component is quantized using a four-level scalar quantizer. Therefore, the number of bits required for each frame is equal to $10 \log_2(4) = 20$ bits. Using the scalar lattice D_{10} quantizer, described in section 3.1, the bit rate of 19 bits/sec is achieved at the cost of a small degradation in performance. Using the lattice quantizer, the sum of the indices at the scalar quantizers is forced to be even. If the sum of the indices is odd, one of the indices must be changed in such a way that a minimum increase in the distortion is achieved. A trellis with two states and 10 stages is employed to achieve this goal. The Viterbi algorithm is used to find the minimum distortion path. As mentioned earlier, the spectral distortion defined by equation (2.14) should be used as the distance measure. Because of the enormous computational complexity associated with the spectral distortion measure, a Euclidean distance measure is used to find the LSF component whose index must be changed.

The computational complexity is 98 flops for the scalar lattice N_0 quantizer, while the computational complexity for the scalar lattice D_{10} quantizer is 178 flops. A flop is defined a floating point addition (subtraction) or multiplication (division). The memory requirements for both cases are the same and is equal to number of bytes necessary to represent 40 floating point numbers. Since quantization is done for each single frame separately, the delay time is equal to 20 ms. It means that the receiver and the transmitter have a delay difference of 20 ms due to this kind of block coding.

4.4 Double Frame Quantization

The idea is similar to single frame quantization, however two subsequent frames are considered as a block. Individual LSF parameters are quantized using a two dimensional vector quantizer. Differential encoding described before is still employed and the same training database is used to design the quantizer.

Each LSF component, now a two dimensional vector, is quantized using a 16 level two dimensional quantizer. The number of bits required for each frame is equal to $10\log_2(16)/2 = 20$ bits. At the expense of a little degradation in performance, using the lattice structure D_{10}, a bit-rate of 19 bits/sec is obtained. A trellis structure similar to the scalar quantizer described in the previous section is employed. The sum of the transmitted indices is forced to be an even number. The Viterbi algorithm is employed and the Euclidean distance measure is used as the distortion measure. The computational complexity is 418 flops for the double frame quantizer, while the computational complexity for the double frame lattice quantizer is 533 flops.

The memory requirements for both cases are identical and is equal to 320 floating point numbers. Since each double frame is considered as a block, the delay time is equal to 40 ms. Mainly at the expense of an extra delay of 20 ms, a major improvement in performance is achieved.

4.5 Switched Quantization

Two successive frames are quantized using the single frame and double frame quantizers, described in the last sections, and the one which achieves lower distortion is chosen. One bit is used as *switch bit* to indicate single frame or

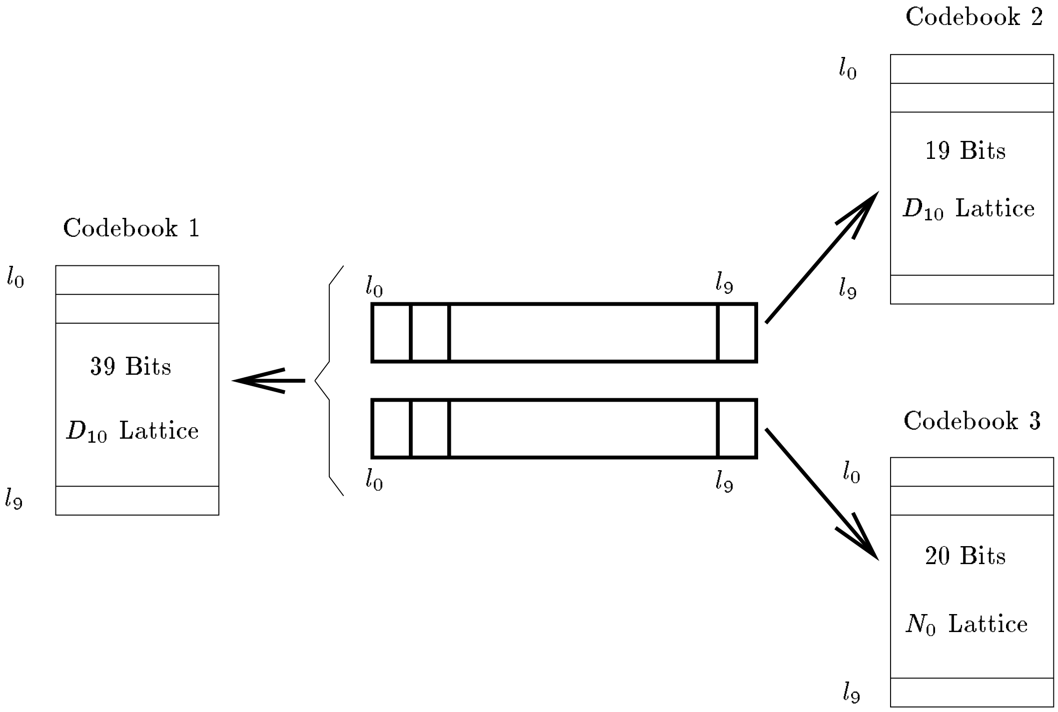

Figure 4.2: Single and Double Frame Codebooks

double frame quantization. If the double frame quantizer has a smaller distance to the original LSF vectors, 39 bits are used for the double frame lattice D_{10} quantizer. Otherwise, the first frame is quantized using the scalar lattice D_{10} quantizer with 19 bits and the second frame is quantized using the scalar lattice N_0 quantizer, with 20 bits. Therefore, the number of bits required to quantize a frame is fixed and equal to 20 bits. Figure 4.2 depicts this method. It is found that, about 20% of the time, single frame quantization achieves lower distortion as compared to the double frame quantization.

Chapter 5

Numerical Results and Performance Evaluation

The split vector quantizer is used as a basis for comparison of the results. The LSF parameter vector is split into two parts. The first part has the first four LSFs and the second part, the remaining six LSFs.The total number of bits available for LSF quantization are divided equally to each part. Therefore, for a 20 bits/frame LSF quantizer, the number of bits allocated for each of the two parts is 10 bits. The weighted Euclidean distance measure discussed in section 3.4 is employed for the design of the split vector quantizer. The Euclidean distance measure is used for the design of the single frame, double frame and switched quantizers. The performance is evaluated on a common database.

5.1 Memory Requirement and Computational Complexity

The amount of static memory (ROM) needed and computational complexity for single frame, double frame, switched quantization and split vector quantization are shown in table 5.1. All the results are for a bit-rate of 20 bits per frame. The unit of memory is *float* which is the number of bytes to represent a floating point number. The complexity is expressed in *flop*, which is the number of necessary floating point additions and multiplications to quantize one LSF vector.

It can be seen from this table that the switched lattice-based quantization has significantly less memory requirement and less computational complexity compared to split vector quantization.

5.2 Spectral Distortion

The spectral distortion measure as presented in equation (2.14) with $f_1 = 60$ Hz and $f_2 = 3500$ Hz is employed to measure the objective quality of the quantized LSF parameters. The results are summarized in table 5.2. It can be seen from the table that the proposed switched quantizer has comparable performance to the split vector quantizer. It can also be seen that switched quantization mainly decreases the percentage of the outliers compared to the

Method	ROM (floats)	Computational Complexity(Flops)
Single Frame	40	98
Double Frame	320	418
Switched	400	672
SplitVQ	10240	38 K

Table 5.1: Memory requirement and computational complexity for a bit-rate of 20 bits per frame

double frame quantizer.

To study the robustness to the channel errors, the results for the channel bit error probability of $P_e = 0$, $P_e = 10^{-2}$, $P_e = 10^{-4}$ and $P_e = 10^{-6}$ are shown in the tables 5.2, 5.3, 5.4 and 5.5, respectively. It is assumed that for the switched quantization, the switch bit is transmitted without error.

Method	Spectral Distortion(dB)	Outliers(Percent)
Single Frame	1.82	29.97
Double Frame	1.38	12.94
Switched	1.32	8.79
SplitVQ	1.26	6.53

Table 5.2: Spectral distortion performance for $P_e = 0$

Method	Spectral Distortion(dB)	Outliers(Percent)
Single Frame	2.27	41.71
Double Frame	1.91	29.75
Switched	1.82	24.80
SplitVQ	2.09	22.95

Table 5.3: Spectral distortion performance for $P_e = 10^{-2}$

Method	Spectral Distortion(dB)	Outliers(Percent)
Single Frame	1.83	30.10
Double Frame	1.39	13.11
Switched	1.32	8.97
SplitVQ	1.27	6.70

Table 5.4: Spectral distortion performance for $P_e = 10^{-4}$

Method	Spectral Distortion(dB)	Outliers(Percent)
Single Frame	1.82	29.98
Double Frame	1.38	12.94
Switched	1.32	8.80
SplitVQ	1.26	6.53

Table 5.5: Spectral distortion performance for $P_e = 10^{-6}$

5.3 Summary

Two successive frames are first quantized using a scalar quantizer, as well as a two dimensional vector quantizer, and subsequently adjusted for a lower bit-rate using a lattice. Lattice-based double frame and single frame quantization is performed for each frame and the one which results in a lower distortion is chosen. One bit is used to indicate single frame or double frame quantization. If the double frame quantizer results in a smaller distance to the original LSF vectors, 39 bits are used for the double frame lattice D_{10} quantizer. Otherwise, the first frame is quantized using the scalar lattice D_{10} quantizer with 19 bits and the second frame is quantized using the scalar lattice N_0 quantizer, with 20 bits. The total number of bits required to quantize a frame is fixed and equal to 20 bits.

A Euclidean distance measure is used as the distortion measure for the design, while a spectral distortion measure[1] is used for the evaluation of the performance. A first-order prediction filter is used to form LSF differences in a closed loop. The parameters for the prediction filter are calculated from a LSF database. The results are compared to the split vector quantizer using a common database for test and training.

[1] Refer to section 2.3.

5.4 Conclusion

In this research paper, an algorithm for the quantization of LSF parameters of speech has been presented. A split vector quantizer[12] is used as a basis for comparison. It has been shown that, the 20 bits/frame switched lattice-based quantizer can achieve the same performance as spilt vector quantizer with very low memory requirements and very low computational complexity. This performance is mainly attributed to the double frame quantization of LSF parameters. Effect of channel errors on the performance of this quantizer has also been investigated. It has been found that the switched lattice-based quantizer is more robust to channel errors than the split vector quantizer.

5.5 Further Work

The proposed quantizer is designed for the fixed rate of 20 bits per frame using uniform bit allocation. To use the quantizer with different bit-rates, it is necessary to allocate a nonuniform number of bits to the individual LSF parameters. Since the first LSF vector components are more important than the other components, more bits should be assigned to them. It could be useful to design the switched lattice-based quantizer using nonuniform bit allocation schemes so that the overall desired bit-rate is achieved.

Bibliography

[1] Y.Linde, A.Buzo and R.M Gray, "An Algorithm for Vector Quantization Design", IEEE Transaction on Communication, vol. COM-28, pp. 84-95, Jan. 1980.

[2] S.P.Loyd, "Least squares quantization in PCM", Bell Lab. Tech. Note (1957). Published in *IEEE Transaction Information Theory,* vol.IT-28, pp.204-206, May 1989.

[3] F. Itakura, "Line Spectrum Representation of Linear Predictive Coefficient of Speech Signals" *Journal Acoustic Soc. Am.,* vol.57, p. 535, April 1975.

[4] G.D.Forney, "The Viterbi Algorithm", *Proc. IEEE,* vol.61, pp.268-278, 1973.

[5] A.Gresho and R.M Gray, *Vector Quantization and Signal Compression,* Kluwer Academic Publishers, 1992.

[6] K.Sayood, *Introduction to Data Compression,* Morgan Kaufmann Publishers, 1996.

[7] K.K. Paliwal and W.B. Kleijn, "Quantization of LPC Parameters", *Speech Coding and Synthesis,* pp.433-466, Elsevier Science B.V., 1995.

[8] N.S. Jayrant and P. Noll, *Digital Coding of Waveforms,* Bell Telephone Laboratories Incorporated, 1984.

[9] F.Soong and B.Juang, "Line Spectral Pairs (LSP) and Speech Data Compression", *Proc. ICASSP,* pp. 1.10.1 - 1.10.4, 1984.

[10] F. Soong and B. Juang, "Optimal Quantization of LSP parameters", *Proc. ICASSP,* pp. 394-397, 1988.

[11] F. Soong and B. Juang, "Optimal Quantization of LSP parameters Using Delayed Decisions", *Proc. ICASSP,* pp. 185-188, 1990.

[12] K.K. Paliwal and B.S. Atal,"Efficient vector quantization of LPC parameters at 24 bits/frame", IEEE Trans. Speech and Audio Processing, vol.1, pp. 3-14, 1993.

[13] F. Soong and B. Juang,"Optimal quantization of lsp parameters", *Proc. Int. Conf. Acoust., Speech, Signal Processing,* pp.394-397, 1988

[14] D.O'Shaughnessy, *Speech communication, Human and Machine,* Addisin-Wesley, Series in Electrical Engineering, 1987.

[15] I.A. Gerson, M.A. Jasink, "Vector Sum Excited Linear Prediction(VSELP) Speech Coding at 8 KBPS", *IEEE International Conference on Acoustics, Speech and Signal Processing,* vol.2970, pp. 461-4 vol.1., 1990.

[16] A.S. Spanias, "Speech Coding: A Tutorial Review", *Proceeding of IEEE,* vol.82, No.10, 1994.

[17] M.R. Schroeder, B.S. Atal, "Speech coding using efficient block codes", *Proc. Intl. Conf. ASSP.,* vol. 3, pp. 1668-1671, May 1982.