Single and Double Frame Quantization of LSF Parameters Using Noise Feed-back Coding*

A. R. Fazel and A. K. Khandani
Dept. of Elec. and Comp. Eng., University of Waterloo, Waterloo, Ont., N2L 3G1
Email: arfazel@shannon.uwaterloo.ca, khandani@shannon.uwaterloo.ca

Abstract: We present a scheme for single frame (20 msec) and double frame (40 msec) quantization of LSF parameters in a CELP speech coder using noise feed back coding. To improve the performance, an appropriate lattice structure is used as the quantizer in the noise feedback loop. We also consider a switched structure based on using either a double frame quantizer or two single frame quantizers for two subsequent frames where an extra bit is used to specify the choice offering a lower distortion. Numerical results are presented showing an excellent performance with very low complexity.

1 Introduction

The most common speech coding schemes currently in use are based on the technique of Code-Exited Linear Prediction (CELP) [2, 3]. These are analysis-by-synthesis Linear Prediction (LP) schemes where the speech signal is represented as the output of a linear time-varying predictor filter, excited by an error signal. Then, a quantized version of the predictor co-efficients, plus the corresponding error signal are transmitted to the receiver side.

Direct quantization of the LP coefficients is usually not done as small quantization errors in the individual coefficients can produce relatively large spectral errors and can also result in instability of the filter. Because of these problems, it is necessary to transform the LP coefficients to another representation which is more appropriate for quantization. The most popular form of such representations is the Line Spectral Frequency (LSF) [4].

In practice, the input signal is partitioned into frames and sub-frames where the excitation signal is updated in each sub-frame (usually 5 msec) and the filter structure is updated in each frame (usually 20 msec). Although a frame size of 20 msec is very common in practice, it is well known that a delay of 40 msec is still well acceptable to human auditory system. As an example, in the TIA IS-641 Enhanced Full Rate Speech Codec (EFRC) which is currently a common standard, the channel interleaving involves two subsequent frames (each of 20 msec) resulting in an overall delay of 40 msec. Our main motivation for the current work is to take advantage of the entire allowable delay of 40 msec to enhance the performance of the LSF quantization scheme.

A spectral distortion measure is commonly used as an objective measure to asses the quality of the LPC quantization [5, 6]. A frame is called an outlier if it has a spectral distortion greater than 2dB. In the design of the quantizer, Euclidean distance measure is usually used as a replacement for the spectral distortion to reduce the complexity. In practice, a weighted Euclidean distance measure which assigns weights to individual LSFs according to their spectral sensitivity is usually employed.

In [5], Soong and Juang have used differential (scalar) quantization of LSFs using LBG algorithm. As a continuation to [5], Soong and Juang in [7] have proposed an algorithm which also incorporates the spectral sensitivities of individual LSFs in the design procedure. Another relevant method is based on using delayed-decision coding [8, 9] to quantize the LSF differences.

Vector quantizers consider the entire set of LPC parameters and allow for direct minimization of quantization distortion. Vector quantizers result in a small distortion, but unfortunately, require a large memory and computational complexity. Due to this fact, a sub-optimal vector quantizer has to be used providing a compromise between complexity and performance. Various forms of suboptimal vector quantizers have been proposed to achieve such tradeoffs [10]. In the same line, a type of product-code vector quantizer, known as Split Vector Quantizer (Split VQ), has been studied in [6]. In this scheme, the vector of LSF values is split into a number of smaller parts and each part is quantized separately using vector quantization. Split VQ is currently the most common method for LPC quantization and is used in many of the standards (including IS-641). This method is implemented and used as the basis for the comparison in this work.

The LSF values within a frame are dependent on each other (*intra-frame* dependency). This dependency can be used to reduce the quantizer bit rate. We make use of the traditional noise feedback method to quantize the prediction error between successive LSF values. This is done using a scalar quantizer, or a two-dimensional vector quantizer. In both cases, a lattice constraint is imposed on the labels of the selected points to further reduce the bit rate.

---

*This work has been supported by Nortel Networks and by Communications and Information Technology Ontario (CITO). A more detailed version of this work can be found in [1].

The differential quantization scheme used in this work is shown in Fig. 1. The prediction coefficients, α_i, β_i, are computed using standard methods of linear prediction by minimizing the average energy of the prediction error, i.e., $e(i+1) = L(i+1) - (\alpha_i L_i + \beta_i)$. The database for the training of the quantizers is used to calculate the constants α_i, β_i.

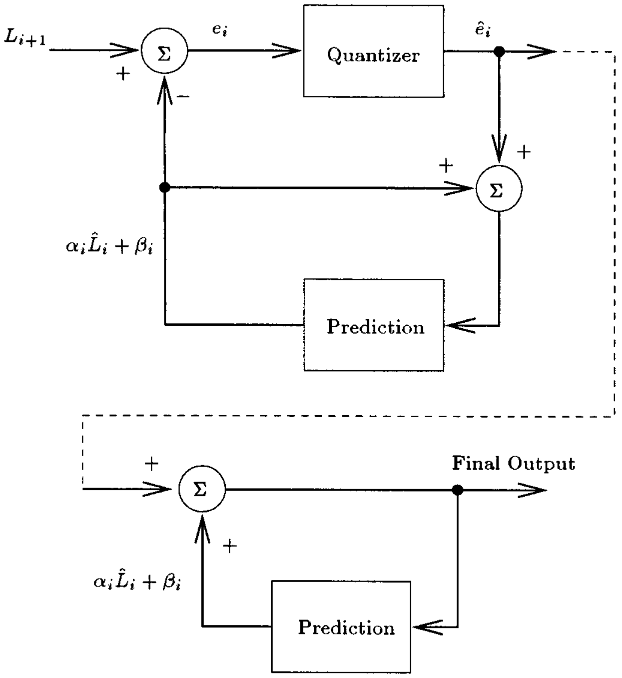

Figure 1: Block diagram of noise feed-back quantization.

The LSF vectors in adjacent frames are also dependent on each other (*inter-frame* dependency). This property can be used to further reduce the bit rate. Since all inter-frame encoders use the information of the previous frames, they suffer from error propagation in communication over noisy channels. In the present work, we limit the scope of such inter-frame prediction to two adjacent frames (40 msec) to limit the effect of error propagation. As already mentioned, in the TIA IS-641 Enhanced Full Rate Speech Codec (EFRC), the channel interleaving involves two subsequent frames (each of 20 msec) and this is fully compatible with the delay characteristics of the method proposed here.

It is well known that the performance of an LSF quantizer largely depends on the training and test databases used (it is also sensitive to the pre-filtering applied to speech). Therefore, it is usually not meaningful to compare the performance of different LSF quantizers unless these are designed and tested using common databases. Due to this fact, a limited comparison is performed in this work. For this purpose, we have implemented a Split VQ which is currently the most common method of LPC quantization.

The database (used for all the methods) contains approximately 175,000 LSF vectors to train the quantizers and 11,000 vectors to test the performance. Weighted Euclidean distance (with the values given in the ITU-T G.729 standard) is used to design the quantizers and spectral distortion is used to evaluate the performance.

2 Lattice Quantization

A real lattice Λ is a discrete set of vectors in real Euclidean N-space, $\mathcal{R}^N$, that forms a group under ordinary vector addition. Generally, an N-D lattice Λ in $\mathcal{R}^N$ has the form $\Lambda = \{\sum_{i=1}^{N} l_i v_i;\ l_i \text{ is an integer}\}$, where $v_1,\ldots,v_N$ are N linearly independent vectors in R^N called the generator vectors for the lattice. Thus a lattice consists of all integral linear combinations of its generator vectors. *Fundamental region* of a lattice is a building block which when repeated many times fills the whole space with just one lattice point in each copy. A lattice automatically partitions $\mathcal{R}^N$ into the collection of nearest neighbors or *Voronoi* regions corresponding to the lattice points. A lattice quantizer is based on using the points of a lattice to partition the space into the quantization regions. In this case, the structure of the lattice is used to facilitate the quantization operation.

For quantization, given a point $\mathbf{r} = (r_1, \ldots, r_N)$, we find a point $\mathbf{x}$ in the lattice which has the minimum (weighted) Euclidean distance to $\mathbf{r}$. The objective is to choose a lattice with a low encoding complexity and with a high quantization gain. In this work, lattice D_{10} is employed which is composed of the points $(x_1, x_2, \ldots, x_{10})$, for which

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

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

The structure of this lattice can be represented by a trellis diagram as shown in Fig. 2 where k_i represents the index of quantization levels for LSF prediction errors and the states correspond to the LSF values. A modification of the Viterbi algorithm is used to find the path with the minimum additive distortion through the trellis. In this case, the survivor path ending to any given state determines the corresponding LSF value through addition of the LSF prediction errors on the branches of that path. This LSF value is also used to perform the prediction for all the branches starting from that state (refer to Fig. 3).

0-7803-7097-1/01/$10.00 ©2001 IEEE

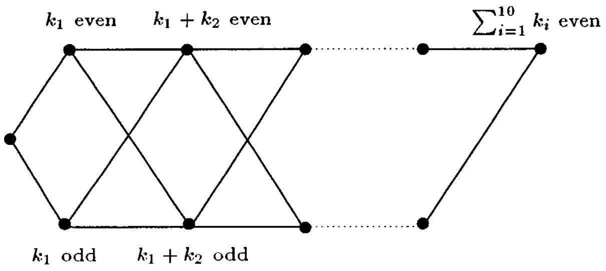

Figure 2: Trellis-based decoding of the lattice D_{10}.

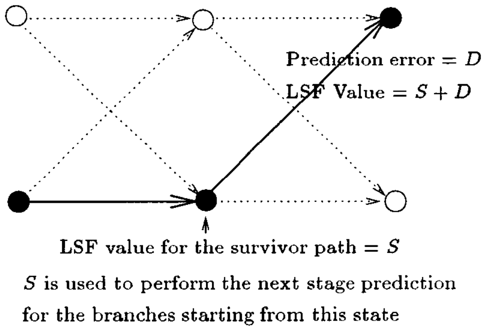

Figure 3: Search and prediction through the trellis.

3 Single Frame Quantization

To exploit the intra-frame correlation of a LSF vector, the differential quantization described in the previous section is employed. In this case, using a lattice quantizer with four points along each dimension results in a quantizer of 19 bits. Note that this scheme results in two parallel branches in each section of the trellis. The computational complexity is 100 flops[1]/frame for the scalar quantization, and 180 flops/frame for the lattice D_{10}. The memory requirement for both the scalar and lattice quantizers are the same and equal to number of bytes used to store 40 floating point values. Since quantization is done for each frame separately, the quantization delay is equal to 20 msec.

4 Double Frame Quantization

The idea is similar to single frame quantization, however, two subsequent frames are quantized together using a set of two dimensional vector quantizers. Differential encoding described before is still employed, but each component, now a two dimensional vector of LSF prediction errors, is quantized using a 16 level two dimensional quantizer. Note that this scheme results in four parallel branches in each section of the trellis. The total number of bits for the two frames is 40. A trellis structure similar to the scalar quantizer described earlier is employed where the sum of the transmitted indices is forced to be an even number (resulting in a bit rate of 39). The computational complexity is 420 flops/frame for the double frame quantizer without lattice, and 530 flops/frame for double frame quantizer with lattice. The memory requirements for both cases are identical and equal to 320 floating point numbers.

5 Switched Quantization

In this scheme, two successive frames are quantized using the single frame and double frame quantizers, described earlier, and the one which achieves a lower distortion is chosen. One bit is used as *switch bit* to indicate single frame or double frame quantization. In the case of two single frames, the first frame is quantized using lattice D_{10} (19 bits) and the second frame is quantized using scalar quantization (20 bits). Therefore, the number of bits required to quantize two frames is fixed and equal to 40 bits. It is found that, about 20% of the time, single frame quantization achieves lower distortion as compared to the double frame quantization.

6 Numerical Results

The split vector quantizer is used as a basis for comparison. The LSF parameter vector is split into two parts of size four and six, respectively. For a 20 bits/frame quantizer, the number of bits allocated to each of the two parts is 10 bits. The amount of static memory (ROM) and computational complexity for single frame, double frame, switched quantization and split vector quantization are shown in Table 1. 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 for a very close performance.

Method	ROM (float/frame)	Computation (flop/frame)
Single Frame	40	100
Double Frame	320	420
Switched	400	670
SplitVQ	10240	38000

Table 1: Memory and computational complexity per frame.

Method	Spectral Distortion(dB)	Outliers(%)
Single Frame	1.8	30.0
Double Frame	1.4	12.9
Switched	1.3	8.8
SplitVQ	1.3	6.5

Table 2: Spectral distortion and outlier for different methods.

---

[1] A flop is defined as a floating point addition (subtraction) or multiplication (division).

7 Summary

An algorithm for the switched lattice-based quantization of LSF parameters is presented. The proposed method achieves a performance very close to spilt vector quantizer with much lower memory requirements and computational complexity.

References

[1] A. R. Fazel, "Switched lattice-based quantization of LSF parameters," M. Eng. Thesis, Dept. of Elec. and Comp. Eng., University of Waterloo, Sept. 1999.

[2] B. Atal, "Predictive coding of speech at low bit rates", *IEEE Trans. Communications*, Vol. COM-30, No. 4, April 1982, pp. 600–606.

[3] M.R. Schroeder and B. Atal, "Code-excited linear prediction (CELP): High quality speech at very low bit rates," *proc. ICASSP-85*, Tampa, Florida, April 1985, pp. 937—941.

[4] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Am.*, vol.57, p. 535, April 1975.

[5] F. Soong and B. Juang, "Line spectrum pair (LSP) and speech data compression," *Proc. Conf. Acoust., Speech, Signal Processing,* (San Diego), pp.1.1O.1-1.1O.4, 1984.

[6] K. K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *J. Acoust. Soc. Am.,* vol. 87, p. 539, 1990 (also see *Proc. Int. Conf Acoust., Speech, Signal Processing,* (Toronto), pp. 661-664, 1991, and *IEEE Trans. Speech and Audio Processing,* vol.1, pp.3-14, 1993).

[7] F. Soong and B. Juang, "Optimal quantization of lsp parameters," *Proc. Con. Acoust. Speech, Signal Processing,* (New York, NY), pp. 394-397, 1988.

[8] F. Soong and B. Juang, "Optimal quantization of LSP parameters using delayed decisions," *Proc. Conf. Acoust., Speech, Signal Processing,* (Albuquerque), pp. 185-188, 1990.

[9] A. Hagen and P. Hedelin, "Low bit-rate spectral coding in CELP, a new LSP method," *Proc. Conf. Acoust., Speech, Signal Processing,* (Albuquerque), pp. 189-192, 1990.

[10] J. Makhoul, S. Houcos, and H. Gish, "Vector quantization in speech coding," *Proc. IEEE,* vol.73, pp. 1551-1 588, 1985.