DIFFERENTIAL INTERPOLATIVE PREDICTION SCALAR QUANTIZATION OF THE LINE SPECTRAL FREQUENCIES FOR LOW BIT-RATE SPECTRAL CODING OF SPEECH

H.R. Sadegh Mohammadi* and W.H. Holmes**

* Electrical Eng. Research Center, Jahad Daneshgahi, IUST University, Iran
** Electrical Engineering School, University of NSW, Australia

ABSTRACT - Line Spectral Frequencies (LSFs) have been used widely as a set of parameters for representation of the all-pole filter in linear prediction based speech coders. In this paper, a new method called Differential Interpolative Prediction Scalar Quantization (DIPSQ) is proposed for coding the LSFs. It is shown by simulation that with the new system of quantization these parameters can be encoded more efficiently than with ordinary scalar quantization as used in the standard CELP coders.

INTRODUCTION

In low bit-rate speech coding methods, preserving the shape of the spectral envelope is quite important. The reason for this is the close relationship between the short-term spectrum and the formants of the speech signals, which are perceptually very significant. In analysis-by-synthesis coding methods the spectral envelope is usually modelled by an all-pole filter. This filter may be characterized in many ways, but line spectral frequencies are the most popular parameters for spectral coding of speech.

With Code-Excited Linear Predictive coding (CELP), which has for example been used as the US Department Of Defense (DoD) standard FS1016 (1991), a tenth order LP filter is commonly used as the all-pole filter. There are thus 10 line spectral frequencies to be quantized. The FS1016 standard uses a fixed quantization scheme with 34 bits/frame, which are allocated unequally between the different parameters (3, 4, 4, 4, 4, 3, 3, 3, 3, 3). Quantization of the LSFs is performed independently, using scalar quantization based on several non-uniform reference levels.

Among different processing techniques that are employed in the development of more efficient quantization methods, particularly scalar quantization, prediction is probably the most powerful and most popular method. Differential quantization and quantization schemes based on interpolation (with their variants) are two common groups of methods that in principle use a prediction technique. On this basis, several methods have been suggested to take advantage of intra-frame and inter-frame correlation between the LSFs. Among the scalar quantization (SQ) schemes, differential (incremental) methods, adaptive methods and/or combinations of these methods with simple SQ and with fixed or adaptive bit allocation have been presented.

In this paper a new method called Differential Interpolative Prediction Scalar Quantization (DIPSQ) of the LSFs is presented, which exploits the intra-frame correlations between the line spectrum pairs. This method uses both interpolation and differentiation to efficiently quantize the LSFs.

This paper is organized as follows. The next section deals with the principles of the proposed method. Then the realization of the method will be discussed and the training of the associated quantization table will be addressed. An experimental simulation based on a CELP speech coder is also described. Finally, conclusions are presented.

DIFFERENTIAL INTERPOLATIVE PREDICTION SCALAR QUANTIZATION

The DIPSQ method combines the ideas of interpolation and differentiation for providing a more efficient quantization scheme. In this method, it is supposed that some of the parameters (LSFs in our case) are scalar quantized independently. These are called *base parameters*. The other parameters (LSFs) are scalar quantized in relation to the base parameters; therefore, they will be known as *dependent parameters*.

Suppose that f_k and f_l ($l > k$) are two adjacent base parameters (LSFs). This means the there is no base parameter between them, but there are some (or at least one) dependent LSF between these two base parameters. The quantized values of these base LSFs are defined as

$$\begin{aligned} \hat{f}_k &= \mathrm{Q_S}(f_k) \\ \hat{f}_l &= \mathrm{Q_S}(f_l) \end{aligned} \quad , \tag{1}$$

where $\mathrm{Q_S}(.)$ represents the independent scalar quantization process. (However, it should be remembered that differing quantization levels are used for the different base parameters.)

Coarse predictions of the other LSFs that are located between the k-th and l-th parameters are calculated by linear interpolations between $\hat{f}_k$ and $\hat{f}_l$, i.e.

$$\tilde{f}_m = \frac{\hat{f}_l - \hat{f}_k}{(l-k)} \times (m-k) + \hat{f}_k \qquad k < m < l \ , \tag{2}$$

where $\tilde{f}_m$ represents the coarse prediction of the m-th LSF. The factor that $(m-k)$ is multiplied by is called the *interpolation step*. In the next stage of the process, the normalized interpolated prediction difference for that parameter is defined as the difference between the actual LSF and its coarse prediction divided by the predicted value; i.e.

$$d_m \stackrel{\Delta}{=} \frac{f_m - \tilde{f}_m}{\tilde{f}_m} \ . \tag{3}$$

Finally, this normalized difference is scalar quantized and the index of the selected quantization level is sent to the decoder, that is

$$\hat{d}_m = \mathrm{Q_S}(d_m) \qquad k < m < l \ . \tag{4}$$

On the decoder side the inverse process is applied for recovering the quantized levels of the LSF parameter. Therefore, the quantized m-th LSF will be calculated as

$$\hat{f}_m = \tilde{f}_m(\hat{d}_m + 1) \ , \tag{5}$$

where $\tilde{f}_m$ is calculated from Equation (2).

Figure 1 shows a typical implementation of LSF quantization by the DIPSQ method for the base and dependent parameters. In the decoder the inverse process is applied for recovering the quantized levels of the LSFs.

Using this method, the quantization noise will be decreased compared to independent scalar quantization (assuming the same total number of bits is used).

REALIZATION AND TRAINING

The DIPSQ method can be implemented for LSF quantization in various ways. The first decision that has to be made is the selection of the number and locations of the base parameters. For instance, if f_1, f_4, f_7 and f_{10} are chosen to be the base parameters, then there will be two other LSFs between each pair of adjacent base parameters.

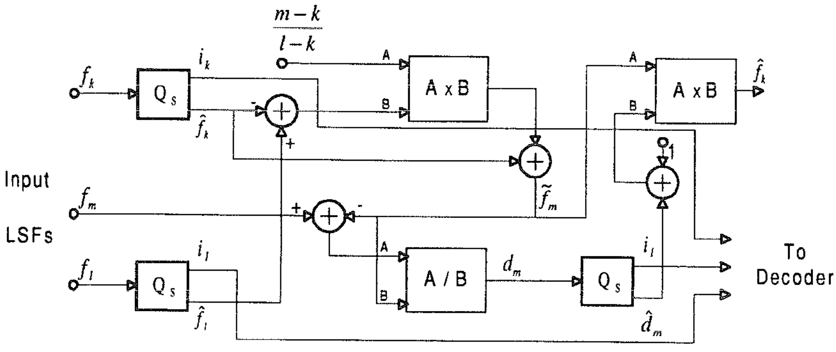

Figure 1. A typical block diagram for DIPSQ quantization of the LSFs

In this study, no attempt has been made to select the best number or locations of the base parameters, since the main goal was the establishment of the fundamentals of the new algorithm. However, an implementation of DIPSQ has been developed for experimental and development purposes. It was decided to choose only the first and the last LSFs (i.e. f_1 and f_{10} for 10th order linear prediction) as the base parameters to be quantized by independent scalar quantization (ISQ).

The quantization levels for these two LSFs could be trained by any proper method, but in our implementation the same quantization table as in the FS1016 standard is used. The remaining LSFs are quantized by dependent scalar quantization as explained earlier. Therefore the coarse predictions of the other LSF frequencies are calculated using the formula

$$\tilde{f}_m = \frac{\hat{f}_{10} - \hat{f}_1}{9} \times (m-1) + \hat{f}_1 \qquad 1 < m < 10\,, \tag{6}$$

and the remaining calculations are performed as above.

A typical example of the coarse prediction process for this experimental implementation is depicted in Figure 2. For each LSF the difference between the exact and predicted values is the vertical distance between the circle and asterisk symbols. In general, as Figure 2 shows, the coarse predictions calculated by this method are good preliminary approximations of the exact LSFs.

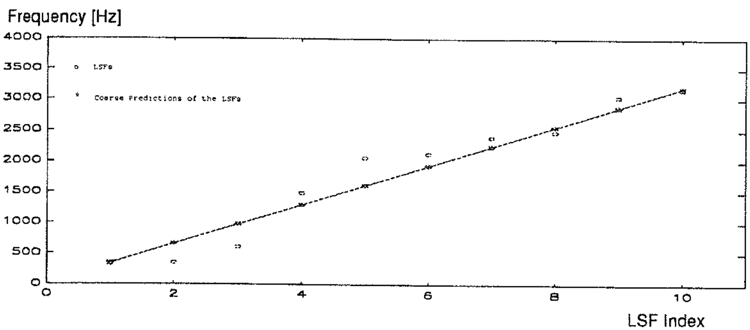

Figure.2. Coarse predictions of the LSFs

The bit allocation for the LSF parameters was chosen to be similar to that of the FS1016 standard. For LSF1 and LSF10 the same quantization levels as in FS1016 have been used (no training), and the other elements of the quantization table are trained by the Lloyd-Max algorithm.

In the first stage of training, LSF1 and LSF10 were independently scalar quantized and the coarse predictions and the normalized interpolative prediction differences were calculated from Equations (2) and (3). Then for each LSF the normalized interpolative prediction difference is used as the training database for the Lloyd-Max algorithm (Lloyd 1982).

COMPUTER SIMULATIONS

We performed an experimental test for the evaluation of the proposed quantization method. A training database which contains 10240 frames of speech signals (each frame equal to 30 ms), spoken by several male and female speakers, was used. The test speech set has 2290 frames of speech. It should be noted that different speakers and sentences are included in the training and test sets. The Lloyd-Max algorithm is used for training the DIPSQ algorithm. The numbers of bits allocated to each LSF were the same as in the FS1016 standard. For comparison purposes we also included an ordinary scalar quantization method with a trained quantization table (prepared by the Lloyd-Max algorithm) in the experiment. The bit allocation for this method was the same as for the others.

The test speech signals were coded with the CELP coder with both the standard and proposed methods as well as with the scalar quantization method with a trained quantization table. Also the same databases have been coded with the CELP coder with unquantized LSFs for comparison. The experimental tests are done using a speech coding simulation computer program written in the C language. This coder is similar to the FS1016 standard speech coder except for the short-term spectrum quantization method.

There is one important point about importing the results of LSF quantization to the simulated speech coding program. That is, regardless of the quantization method used for the spectral coding, the quantized LSFs are always sorted in each frame before handing over to the speech coding program. This retains the ordering property of the LSFs that is necessary for the implementation of the inverse-filter.

Three objective distortion measures are considered for evaluation of the results: spectral distortion (SD), Synthesized Spectral Distortion (SSD) and Segmental Signal to Noise Ratio. The same definitions as those used in Sadegh Mohammadi & Holmes (1994) are considered for these objective measures. The results of quality assessments for our experiment are presented in Table 1.

Table 1. Results of spectral distortion measurements for different quantization methods

Quantization Method	No. of Bits per Frame	Spectral Distortion [dB]	Synthesized Spectral Distortion [dB]	Segmental Signal-to-Noise Ratio [dB]
Unquantized	Infinity	0.00	2.12	9.17
SQ (FS1016 Std.)	34	1.49	2.31	8.73
SQ (Trained)	34	1.25	2.25	8.67
DIPSQ	34	1.19	2.30	8.79

These objective quality assessment results of the quantized short-term speech spectrums and reconstructed speech signals with different distortion measures show that the proposed method provides better quantization results with lower spectral distortion. These results are confirmed through informal subjective tests. The additional computational cost of the DIPSQ method compared to the ordinary scalar quantization method is quite insignificant and it is still much lower than that of fast vector quantization methods.

Note that in this experiment no attempt was made to find the best selection of base and dependent LSFs, since in this research the establishment of the new quantization method was the main goal and not the optimum selection of the parameters. For achieving the best performance with the DIPSQ method it will be necessary to investigate the optimum number and location of the base and dependent parameters and to train the quantization tables for all LSFs on this basis.

CONCLUSION

A new scalar quantization scheme for coding LSFs, called differential interpolative prediction scalar quantization (DIPSQ), is proposed in this paper and its principles are described. The application of this quantization method to spectral coding of speech signals is presented as a case study. The performance of DIPSQ has been compared to ordinary scalar quantization methods. The results show the obvious advantages of DIPSQ over simple scalar quantization, even without optimization of the new method.

REFERENCES

Federal standard 1016 (1991) *Telecommunications: analog to digital conversion of radio voice by 4,800 bit/second code excited linear prediction (CELP),* (National Communications System, Office of Technology and Standards, Washington, DC20305-2010).

Lloyd, S.P. (1982) "Least squares quantization in PCM", *IEEE Trans. Inform. Theory,* 28, pp. 129-137.

Sadegh Mohammadi, H.R. & Holmes, W.H. (1994) "Fine-Coarse Split Vector Quantization: An Efficient Method For Spectral Coding", *Proc. of Fifth Australian Intern. Conf. on Speech Sci. and Tech.*, pp. 118-123, Dec. 1994.