LSF QUANTISATION FOR PITCH SYNCHRONOUS SPEECH CODERS

C. Sturt, S. Villette, A. M. Kondoz

Centre for Communication Systems Research
University of Surrey, Guildford, Surrey, GU2 7XH United Kingdom
Email: c.sturt@eim.surrey.ac.uk

ABSTRACT

The quality of low bit-rate speech coders is reduced at transitions where speech spectral characteristics vary significantly, as usual speech parameter interpolation assumptions fail to correctly model such variations. This paper presents a joint quantisation-interpolation algorithm for coding of LPC parameters in pitch synchronous speech coders to model the rapidly evolving parameters. In this technique a number of sets of pitch synchronous LPC parameters, corresponding to a frame of speech, are jointly coded by coding two reference sets of LSF's and an interpolation trajectory. Coding an interpolation function allows the parameters to vary within the set. The proposed joint quantisation-interpolation coding of the pitch synchronous LSF is evaluated by comparison with time synchronous extraction and linear interpolation. It is also compared with linear interpolation between sets of pitch synchronous LSF's. Comparison results show that the joint quantisation-interpolation method reduces the average spectral distortion when compared to fixed interpolation. The proposed quantiser was included in to the PS-SBLPC coder and informal listening tests carried out. The synthesised speech was found to be of better quality when joint quantisation-interpolation is used.

1. INTRODUCTION

Recently there has been an increased interest in Pitch Synchronous (PS) coding of speech. Recent studies have addressed the design of PS-CELP [1], PS-Multi-Band [2] and PS Sinusoidal Coding [3]. These methods have become popular as classic Time-Synchronous (TS) methods fail to sufficiently exploit many of the properties of the speech signal. PS based coders analyse a speech signal as individual pitch cycles rather than as conventional frames of fixed length. Initially the pitch cycles mush be located and then the short cycles of speech are analysed. PS-LPC analysis is carried out by analysing three cycles of speech, centred on the current coded cycle, at a time. As the LPC analysis is carried out at a finer interval than standard TS-LPC analysis, the extracted coefficients contain higher correlation between neighbouring pitch cycles. The resultant residual, formed by localised inverse filtering, also shows higher correlations [1].

PS analysis of speech signals produces a variable number of LPC parameters per 20ms frame as the number of cycles depend on the pitch of the speech signal. Individual quantisation of the PS parameters would lead to a variable rate coder, which is undesirable in most cases. Therefore it is necessary to include a pitch synchronous to time synchronous conversion to allow for fixed rate quantisation of the LPC parameters. The coding scheme must allow for fixed rate coding of the LPC parameters, whilst at the same time representing parameter variations within the 20ms frame. Guerchi and Mermelstein [1] proposed a joint quantisation-interpolation method to code PS-LPC parameters, assuming linear variations of the LSF parameters within a speech frame. This method codes a set of LSF parameters per frame to minimise the total spectral distortion between the PS extracted LSF parameters and a sequence of decoded vectors assuming linear interpolation. This method does not allow for non-linear evolution of parameters within the frame of speech, which causes quality degradation. The goal of the work presented in this paper is to design a fixed rate quantisation scheme to code PS-LPC parameters for use in pitch synchronous coders, specifically PS-SBLPC [3]. The Joint Quantisation-Interpolation (JQI) algorithm must allow for non-linear parameter evolution within a 20ms frame. A total of 36 bits are available for the quantisation of the LPC parameters in the PS-SBLPC coder operating at 4kbss.

2. JOINT LSF QUANTISATION

It is proposed that a JQI technique be used to encode sets of successive LPC parameters extracted from within a pitch synchronous speech coder. Guerchi and Mermelstein proposed a method in [4] to quantise PS-LPC parameters, assuming linear interpolation of the parameters within a frame. Figure 1a shows the operation of this method. Parameters x_{-1}, x_1 and x_2, are selected so as to minimise the overall distortion between the PS computed parameters and the decoded parameters assuming linear interpolation

within a speech frame. Parameters x_{-1} are from the previous frame and are fixed. Evaluation of this method within the PS-SBLPC coder showed that this method was not sufficient to represent the evolution of LSF parameters over speech transitions. It was found that the LSF parameters did not normally vary linearly within a frame, and that the resultant synthesised speech lacked sharpness at such places. In order to account for this, it is proposed to allow for non-linear evolution of LSF parameters within the frame. Figure 1b shows the proposed method, showing a stepped interpolation. In order to allow for greater variation within the frame, the PS-LSF's are represented as a weighted combination of three sets of LSF's, two from the current frame and one from the previous frame.

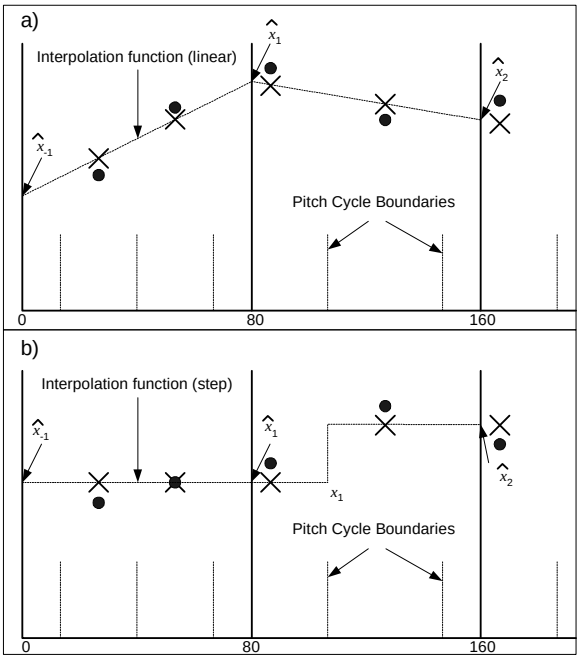

Figure 1: Illustration of the combined interpolation-quantisation process. ● extracted parameters, × quantised parameters, x_{-1},x_1,x_2 target vector.

In the JQI quantiser, the LPC parameters are coded in the LSF domain, as direct LPC interpolation is not practical. The JQI quantises K sets of 10^{th} order LSF parameters, by jointly quantising 2 sets of LSF parameters and an evolution function. The evolution function is a combination of γ times the last LSF from the previous frame, α times the first quantised LSF and β times the second quantised LSF from the current frame. α,β and γ are functions of K, the pitch cycles within a frame. In order to restrict the range of the quantised PSL-LSF's the values of the evolution function are constrained such that $\alpha_k+\beta_k+\gamma_k=1$. Hence $\gamma_k=1-\alpha_k-\beta_k$. In order to reduce the number of bits required, the two interpolation functions α and β are jointly quantised by concatenating to form one single vector. The synthesised LSFs Y_k take the form of:

$$Y_k^j = (1-\alpha_k-\beta_k)\hat{x}_{-1}^j + \alpha_k \hat{x}_1^j + \beta_k \hat{x}_2^j$$

where $\hat{x}_1$ and $\hat{x}_2$ are the two dequantised sets of LSF's of the current frame, and $\hat{x}_{-1}$ is the final set of dequantised LSF's from the previous frame. The quantisation error E is given by:

$$E = \sum_{k=0}^{K}\left(\sum_{j=0}^{9}\left(\lambda_k^j - Y_k^j\right)^2\right)$$

λ_k are the sets of K extracted LSF's.

By taking:

$$\frac{\partial E}{\partial \alpha_k} = 0 \text{ and } \frac{\partial E}{\partial \beta_k} = 0$$

and solving for α_k and β_k we obtain the optimum values to minimise E given $\hat{x}_1$ and $\hat{x}_2$. We have:

$$\alpha_k = \frac{\sum_{j=0}^{9} a^j c_k^j \cdot \sum_{j=0}^{9} a^j b^j - \sum_{j=0}^{9} b^j c_k^j \cdot \sum_{j=0}^{9}\left(a^j\right)^2}{\left(\sum_{j=0}^{9} a^j b^j\right)^2 - \sum_{j=0}^{9}\left(b^j\right)^2 \cdot \sum_{j=0}^{9}\left(a^j\right)^2}$$

$$\beta_k = \frac{\sum_{ji=0}^{9} b^k c_k^j \cdot \sum_{j=0}^{9} b^j a^j - \sum_{j=0}^{9} a^j c_k^j \cdot \sum_{j=0}^{9}\left(b^j\right)^2}{\left(\sum_{j=0}^{9} b^j a^j\right)^2 - \sum_{j=0}^{9}\left(a^j\right)^2 \cdot \sum_{j=0}^{9}\left(b^j\right)^2}$$

Where:

$$a^j = x_2^j - x_{-1}^j \qquad b_i = x_1^j - x_{-1}^j \qquad c_{i,k} = \lambda_k^j - x_{-1}^j$$

Hence we have the optimum interpolation function, given $\hat{x}_{-1}, \hat{x}_1$ and $\hat{x}_2$. The quantisation error is also a function of $\hat{x}_1$ and $\hat{x}_2$. Therefore the correct selection of x_1 and x_2, the two sets of parameters to be quantised, affects the quantisation error. Initially, x_1, x_2 can be set to the first and last target vectors (λ_0^j and λ_{K-1}^j) or a set of LSF's extracted over each of the two half of the speech frame. In order to minimise E further, the optimum values of x_1, x_2 can be calculated using the calculated values of α_k and β_k.

By taking:

$$E^j = \sum_{k=0}^{K-1}\left(\lambda_k^j - Y_k^j\right)^2$$

$$E^j = \sum_{k=0}^{K-1} \left(\lambda_k^j - \left((1-\alpha_k-\beta_k)\hat{x}_{-1}^j + \alpha_k \hat{x}_1^j + \beta_k \hat{x}_2^j \right) \right)^2$$

And minimising with respect to each set of LSF by setting:

$$\frac{\partial E^j}{\partial x_1^j} = 0 \qquad \frac{\partial E^j}{\partial x_2^j} = 0$$

and solving for x_1^j , x_2^j giving:

$$x_2^j = \frac{v^j r - x_{-1}^j \mathrm{B} r + x_{-1}^j s r - t u^j + t x_{-1}^j \mathrm{A} - x_{-1}^j t^2}{sr - t^2}$$

$$x_1^j = \frac{u^j - x_{-1}^j \mathrm{A} + x_{-1}^j r + x_{-1}^j t - x_2^j t}{r}$$

Where:

$$\mathrm{A} = \sum_{k=0}^{K-1} \hat{\alpha}_k \qquad r = \sum_{k=0}^{K-1} (\hat{\alpha}_k)^2 \qquad u^j = \sum_{k=0}^{K-1} \hat{\alpha}_k \lambda_k^j$$

$$\mathrm{B} = \sum_{k=0}^{K-1} \hat{\beta}_k \qquad s = \sum_{k=0}^{K-1} (\hat{\beta}_k)^2 \qquad v^j = \sum_{k=0}^{K-1} \hat{\beta}_k \lambda_k^j$$

$$t = \sum_{k=0}^{K-1} \hat{\alpha}_k \hat{\beta}_k$$

By employing an iterative calculation of α_k ,β_k, $\hat{x}_1$ and $\hat{x}_2$ the optimum pairing of interpolation evolution and LSF parameters is found. Figure 2 1shows the operation of the JQI quantiser. It clearly shows that the LSF value does not vary linearly within the frame. Therefore simple linear interpolation would not suffice.

3. QUANTISER DESIGN

The JQI scheme proposed requires the quantisation of two sets of LSF parameters and two evolution vectors, of length K. The two sets of LSF parameters are quantised using a moving average multi-stage joint vector quantiser. The evolution vectors α and β are jointly quantised as a single vector of length 2K. The evolution vector is quantised using separate vector codebook for each of the possible values of K. No evolution information needs to be transmitter when only two pitch cycles are present in the frame, so only nine (3-11 cycles) codebooks are required.

Training vectors to generate the evolution codebooks were produced as follows. Twenty-five minutes of speech from the NTT training database were processed through the PS-SBLPC encoder and PS-LPC parameters extracted. For each frame the first and last set of LPC parameters were quantised with the joint LSF quantiser. The dequantised vectors were then used to calculate the optimum evolution vector αβ. These optimum evolution vectors are then written to the appropriate training database that contains the corresponding vector size of 2K. The nine evolution codebooks were then trained using the LBG algorithm with the separate training databases.

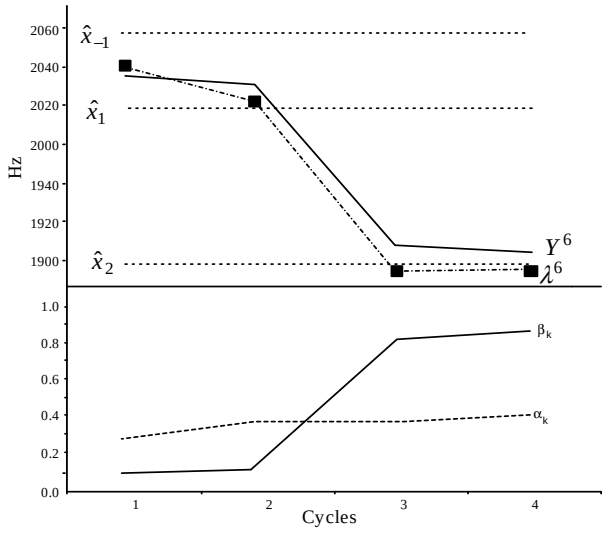

Figure 2, showing the evolution of the 6th LSF within a frame (top), and the resultant quantised evolution vectors α and β (bottom).λ^6 is the extracted 6th LSF, Y^6 the dequantised 6th LSF.

Evolution codebooks containing, sixteen, thirty-two and sixty-four entries were trained. Joint LSF vector quantisers of 30 or 36 bits were used to quantise the LSF's. The resultant log spectral distortion between the PS extracted sets and the dequantised sets of LPC's was calculated for each codebook size. As a reference, the JQI performance was compared with the existing method of time synchronous LPC extraction and linear interpolation. In this case the LSF's were quantised with a 36-bit LSF joint vector quantiser. The 4,5 and 6 bit evolution codebooks were used in conjunction with a 30-bit LSF joint vector quantiser to generate an overall system with 34,35 or 36 bits. Table 1 shows the results obtained.

Code book Size (bits)			Spectral Distortion (dB)
LSF JVQ	Evolution	Total	
36	0	36	1.46
30	4	34	1.24
30	5	35	1.22
30	6	36	1.21

Table 1: Spectral distortion performance of interpolation compared to evolution quantisation.

Although the average spectral distortion of the 34-bit quantiser is lower than that of the 36-bit time synchronous method, listening tests and examination of the resultant LPC filter spectrums showed that with only a four or five bit evolution codebook, there were a significant number of outliers causing distortion. Therefore the final quantiser design includes a 6-bit evolution quantiser and a 30-bit joint vector quantiser. Figure 3 shows four LPC filter

spectra taken from a frame of speech. The top waveform is the PS extracted LPC spectra, the middle shows the dequantised PS-JQI LPC spectra and the final shows the Time Synchronous Linear interpolated LPC spectra. One of the problems of fixed interpolation can be seen in cycle two. An extra peak is still present in this cycle, where as in the original, and in the PS-JQI this peak has disappeared.

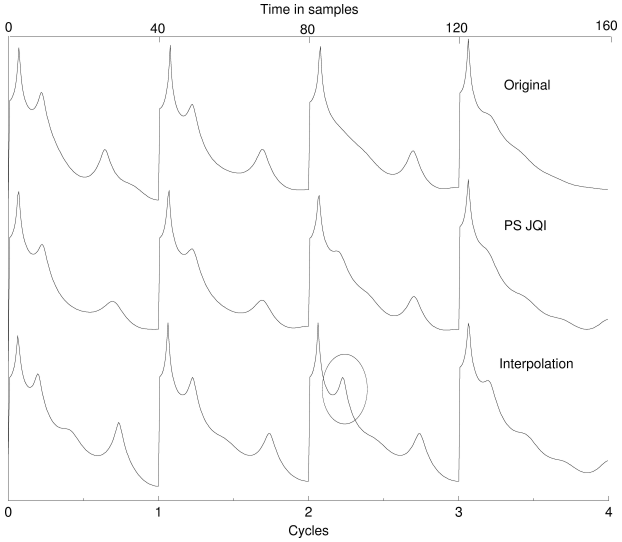

Figure 3: Sets of LPC filter spectrums (log domain). Original, PS JQI and TS Interpolation.

4. EVALUATION AND RESULTS

The JQI of the PS-LPC parameters was evaluated by comparison with standard techniques used to quantise LPC parameters in PS coders. Many coders, including the one presented in [5] use PS synthesis of speech, and generate the PS-LPC parameters by interpolating between two TS-LPC sets extracted at the encoder. The coder in [5] uses energy based interpolation to best generate the PS vectors. Other coders employ linear interpolation. It was found that the results for energy based interpolation and linear interpolation were identical in terms of spectral distortion. 36- and 42-bit joint vector quantisers were used to quantise the TS-LSF parameters in these quantisers.

The JQI algorithm was also compared with linear interpolated vectors generated by interpolating between two optimised sets of parameters as proposed in [4]. 36- and 42-bit joint vector quantisers were used to quantise the optimised parameters in these quantisers. Two JQI quantisers were evaluated, a 36-bit version using a 30-bit joint VQ and a 6-bit evolution quantisation, as well as a 42-bit version using a 36-bit joint VQ and a 6-bit evolution quantisation.

The average log spectral distortion between the extracted and dequantised LPC filter spectrum was calculated in each case, using eight test sentences from different speakers, four male, four female. Table 2 shows the results obtained for the various combinations and bit rates.

Dequantised PS-LSF's from the JQI quantiser were inserted into the PS-SBLPC model. The synthesised speech was compared with that generated using LSF's synthesised by linear interpolation and quantised at the same total bit rate. Informal listening tests showed that the speech quality was improved by using the PS evolution quantiser. The improvements were most evident at speech transitions where the speech changes rapidly.

Interpolation	Average SD (dB)	
	36 bits	42 bits
JQI Evolution	1.21	1.05
PS Linear	1.46	1.41
TS Interpolation	1.68	1.62

Table 2: Comparison of the various quantisation techniques.

5.CONCLUSION

We have presented a new method for the quantisation of PS-LPC parameters. We have compared the performance of the new quantiser with existing techniques at the same total bit rate and have shown that a significant performance gain can be achieved by the introduction of a new LSF evolution function. The technique has been applied to the PS-SBLPC coder and informal listening tests carried out. The results show that speech quality is improved in areas where the characteristics of the speech vary significantly, such as onsets, offsets and transitions.

6. REFERENCES

[1] Guerchi, Y. Qian P. Mermelstein, "Pitch-synchronous linear-prediction analysis by synthesis with reduced pulse densities", In Proceedings of IEEE Workshop on Speech Coding 2000.

[2] H. Yang, S.N. Koh, P. Sivaprakasapillai "Pitch Synchronous Multi-Band (PSMB) Speech Coding". In Proceedings ICASSP-95.

[3] C. Sturt, S. Villette, and A.M. Kondoz, "Pitch Synchronous Split-Band LPC (PS-SBLPC) Vocoder," *Proceedings of IEEE Workshop on Speech Coding 2002, Ibaraki, JAPAN*, pp. 132-134, October 2002.

[4] Guerchi, P. Mermelstein, "Low-rate quantisation of spectral information in a 4kb/s pitch-synchronous CELP coder", In Proceedings ICASSP-2000. Istanbul, Turkey, 2000

[5] S. Villette, K.T Al-Naimi, C. Sturt, A.M. Kondoz and H. Palaz "A 2.4/1.2 kbps SB_LPC Based Speech Coder: The Turkish NATO STANAG Candidate," *Proceedings of IEEE Workshop on Speech Coding 2002, Ibaraki, JAPAN*, pp. 87-89, October 2002.