EFFICIENT CODING OF LSP PARAMETERS USING SPLIT MATRIX QUANTISATION

Professor C.S. Xydeas and C. Papanastasiou

Speech Processing Research Laboratory
Electrical Engineering Division
School of Engineering
University of Manchester
Manchester M13 9PL
UNITED KINGDOM

ABSTRACT

This paper presents a new and efficient LPC quantisation scheme called Split Matrix Quantisation (SMQ). The proposed method can be viewed as an extension of the conventional Split Vector Quantisation process. It operates over N consecutive LPC frames and effectively divides a p×N LSP matrix into K submatrices which are then vector quantised independently. SMQ exploits the interframe redundancy that exists between consecutive sets of LSP coefficients and achieves "transparent" quantisation at 900bits/sec. "High quality" LSP quantisation can be easily obtained at 750bits/sec. These bit rates are based in a 20 msec LPC analysis frame size. Furthermore, SMQ is characterised by relatively low complexity and low storage requirements.

1. INTRODUCTION

The worldwide introduction of new voice communication systems and services has fuelled, in recent years, a significant amount of research activity into low bit rate, high recovered speech quality, coders. Emphasis is now placed on the development of the next generation high speech quality "Source-Filter" vocoding systems capable of operating in the region of 1.2 to 3.2 Kbits/sec. Within this context of low bit rate coding, the efficient quantisation of "filter" parameters in general and LPC filter coefficients in particular, is a necessity in order to release additional bits which can then be allocated to the "excitation source" part of the system.

"Transparent" scalar quantisation of LPC filter coefficients, requires typically 38 to 40 bits per analysis frame [1]. The exploitation of intraframe correlation using differential coding and frequency delayed coding techniques, reduces the above figures to about 30 bits/frame [2]. Lower bit rates can be achieved using Vector Quantisation (VQ). Split-VQ [3] or Single Stage VQ [4] offer the required performance with realistic storage and codebook search characteristics at 24 and 20 bits/frame respectively. Further compression can be obtained in principle by exploiting interframe correlation between sets of LPC coefficients. In this way, adaptive codebook VQ systems have been proposed [5] and combined in certain cases with differential coding and fixed codebooks. Alternatively, Switched Adaptive Interframe Vector Prediction [6,7] can be employed which offers high LPC coefficient quantisation performance at 19 to 21 bits/frame.

Whereas the above schemes attempt to reduce interframe correlation in a backwards manner using past information, Matrix Quantisation [8] allows the introduction of delay into the process and operates simultaneously on sets of p filter coefficients obtained from N successive frames, using VQ principles. Matrix Quantisation has been applied to vocoding systems operating at or below 800 bits/sec [9], where "transparency" in LPC parameter quantisation was not required. In addition, excessive codebook storage and search requirements have been identified with this technique. High complexity and large storage requirements are also prominent in the Joint Segmentation and Quantisation (JSQ) approach [10], which combines optimally a variable bit rate (segmentation) operation and matrix quantisation. This method offers reasonable filter coefficient quantisation performance at about 200 bits/sec.

Although, theoretically, variable rate Joint Segmentation and Quantisation performs better than Matrix Quantisation, the later continues to be of interest because it results in fixed bit rate systems. The inherent Matrix Quantisation drawbacks, i.e. high complexity and large storage requirements, can be solved by splitting the p×N LPC matrix into K submatrices, which are then quantised independently. Within this framework, the paper examines in a comprehensive manner the following four important issues: i) possible representations of the matrix elements, as derived from LSP coefficients, ii) distortion measures and associated time/spectral domain weighting functions used in the codebook design and quantisation processes, iii) objective performance evaluation metrics, which correlate well with subjective experiments performed using synthesised speech, and iv) complexity and codebook storage characteristics. In each case, a number of possible solutions are proposed in a way which leads into several Split Matrix Quantisation designs with different performance/complexity characteristics.

In addition, the paper establishes guidelines for achieving LPC quantisation "transparency", in terms of a new subjectively meaningful objective performance measure. This is because conventional spectral distortion measures, used in conjunction with "single frame" VQ methods [3], can not reflect adequately the subjective importance attached to the "smooth evolution" with time of the short-term envelope information of the

0-7803-2431-5/95 $4.00 © 1995 IEEE

magnitude speech spectra. This spectral behaviour is inherently linked to multiframe VQ schemes in general and Matrix Quantisation in particular.

2. SPLIT MATRIX QUANTISATION

Consider that LPC analysis is applied to speech frames of M msec duration to yield coefficient vectors $\underline{a}(n)=[a_1^n,a_2^n,a_3^n,...,a_p^n]$, where p is the order of the LPC filter and n is the current frame. $\underline{a}(n)$ is then transformed to an LSP representation $\underline{l}(n)=[l_1^n,l_2^n,l_3^n,...,l_p^n]$ and this process, when performed over N consecutive speech frames, provides a pxN LSP matrix.

$$\underline{X}(n)=\begin{bmatrix} l_p^n & l_p^{n+1} & l_p^{n+2} & \cdots & l_p^{n+N-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ l_3^n & l_3^{n+1} & l_3^{n+2} & \cdots & l_3^{n+N-1} \\ l_2^n & l_2^{n+1} & l_2^{n+2} & \cdots & l_2^{n+N-1} \\ l_1^n & l_1^{n+1} & l_1^{n+2} & \cdots & l_1^{n+N-1} \end{bmatrix} \quad (1)$$

In general the above matrix can be split up in to K submatrices:

$$\underline{L}_k(n)=\begin{bmatrix} l_{S(k)-1}^n & l_{S(k)-1}^{n+1} & \cdots & l_{S(k-1)}^{n+N-1} \\ \cdot & \cdot & \cdots & \cdot \\ l_{S(k-1)}^n & l_{S(k-1)}^{n+1} & \cdots & l_{S(k-1)}^{n+N-1} \end{bmatrix} \quad k=1,...,K \quad (2)$$

Note that: $S(k)=\sum_{j=0}^{k} m(j)$, $m(0)=1$, $\sum_{k=1}^{K} m(k)=p$ and

$$\underline{X}(n)=\left[\underline{L}_K(n) \quad \underline{L}_{K-1}(n) \quad \cdots \quad \underline{L}_1(n)\right]^T$$

Now, each row (or set of m(k) rows) in $\underline{X}(n)$ correspond to a "trajectory" in time of spectral coefficients over N successive frames and these trajectories can be vector quantised independently.

In designing the corresponding trajectory codebooks, sequences of $\{\underline{L}_k(n)\}$ submatrices are obtained by sliding a N-frame window, one frame at a time, along the entire training sequence of LPC speech frames. This sliding block technique maximises the number of vectors employed in the codebook design process and ensures that all phoneme transitions present in the input training sequences are captured. Furthermore, in order to maximise SMQ efficiency, different codebook training sequences are generated and hence different codebooks are designed for each of the following three cases a) All N LPC frames are voiced, b) all N LPC frames are unvoiced and c) the N LPC frames segment includes both voiced and unvoiced frames.

Two important issues associated with the SMQ design and operation are discussed in the following sections. These are i) possible representations of the $\underline{X}(n)$ matrix elements and ii) the distortion measure to be employed in the SMQ codebooks design and search processes.

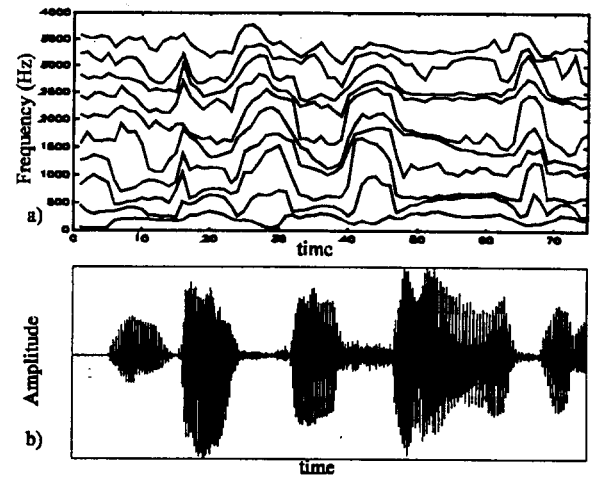

Figure 1 : a) LSP trajectories and b) corresponding speech waveform.

2.1 Representation of matrix elements.

In order to exploit interframe correlation, the pxN matrix elements should reflect the slowly changing with time characteristics of the speech short-term magnitude spectral envelope. In this way, a first intuitive approach can be to employ a formant-bandwidth LSP based representation. Using statistical observations we attempted to relate LSPs to formants and bandwidths by means of a centre frequency (i.e the mean frequency of an LSP pair) and an offset frequency (i.e. half the difference frequency of an LSP pair), as proposed in [11,12]. However, formant/bandwidth information will not always provide smooth trajectories over time and can be therefore difficult to quantise within the SMQ framework. On the other hand, LSPs offer an efficient LPC representation due to their monotonicity property and their relatively smooth evolution over time. Figure 1 illustrates the point made on "smooth" LSP trajectories which are obtained during voiced speech. Both the direct LSP and the Mean-Difference LSP representations have been employed in our SMQ investigations and computer simulation results highlighted the superiority of SMQ schemes based directly on LSP parameters.

2.2 Weighted Distortion Measure.

A weighted Euclidean distortion measure is used in the direct LSP based SMQ codebook design and search processes. This is defined as:

$$D(\underline{L}_k(n),\underline{L}'_k(n))=\sum_{s=0}^{m(k)-1}\left[\sum_{t=0}^{N-1}(LSP_{S(k-1)+s}^{n+t}-LSP'^{n+t}_{S(k-1)+s})^2 w_s(s,t)^2 w_t(t)^2\right] \quad (3)$$

where $\underline{L}'_k(n)$ represents the kth quantised submatrix. A weighting factor $w_t(t)$, which is proportional to the energy and

the degree of voicing in each LPC speech frame, is assigned to all the LSP spectral parameters of that frame. In addition a weighting factor $w_s(s,t)$ is used in Equation 3, which is proportional to the value of the short term power spectrum measured at each frequency associated with the LSP elements of the m(k)×N $\underline{L}_k(n)$ submatrix. $w_s(s,t)$ ensures that distortion associated to high energy spectral areas is emphasised, as compared to low energy spectral regions. In a similar way, $w_t(t)$ ensures that distortion associated to voiced frames is emphasised and thus quantisation accuracy increases in the case of voiced speech segments.

3. LSP QUANTISATION EVALUATION METHODS

The performance of an LPC/LSP quantisation process can be measured in terms of subjective tests and/or objective distortion related measures. Subjective tests are often performed using the arrangement shown in Figure 2. Here, the actual residual signal is used to excite the corresponding LPC filter whose coefficients are quantised. The term "transparent" LPC quantisation refers to the case where, as a result of the noise introduced by quantising the LPC coefficients, no audible distortion can be detected on the $\hat{x}_n(i)$ output signal. Traditionally, objective measures that are used to assess the performance of quantisation schemes operating on LPC parameters, are Spectral Distortion Measure (SDM) variants. SDM is defined as the root mean square difference formed between the original log-power LPC spectrum and the corresponding quantised log-power LPC spectrum. However, these SDM based measures focus on the accuracy of the quantisation process to represent individual LPC frames and thus, fail to capture the perceptually important smooth evolution of LSP parameters across frames. The latter is exploited by SMQ and as a consequence SDM measures do not relate well to SMQ subjective tests.

In pursue of a more accurate measure, we employed a time domain Segmental SNR metric, that is formed using the original $x_n(i)$ and synthesised $\hat{x}_n(i)$ signals (see figure 2). However, the $x_n(i)$ and $\hat{x}_n(i)$ signals are logarithmically (μ-law) processed [13]. This effectively provides a 3.5dB amplification of high frequency spectral components. Furthermore, a weighting factor *Weig(n)* is also used in the Logarithmic Segmental SNR (LogSegSNR) averaging process, which increases the "contribution" of voiced speech frames.

$$Weig_t(n) = [En(n)]^{0.1} \times C \qquad (4)$$

E(n) is the energy of the nth frame and *C*=1 for a voiced frame or *C*=0.01 in the case of an unvoiced frame.

Extensive objective/subjective tests highlighted clearly the perceptual relevance of the LogSegSNR metric. However, it was deemed necessity to combine both the LogSegSNR and average SDM measures in establishing accurate objective performance rules for "transparent" and "high quality" quantisation of LPC parameters. The term "high quality" LPC quantisation indicates

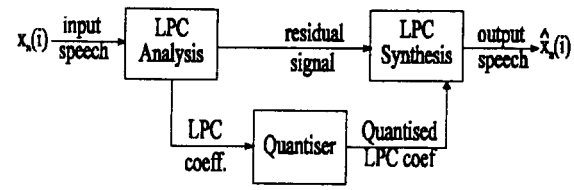

Figure 2: Subjective evaluation system

that although a small difference can be perceived between the input and synthesised signals, nevertheless the effect of LPC quantisation on the quality of the output signal is negligible. In this way, "transparent" LPC quantisation is achieved when LogSegSNR>10dB and AverSDM measured (using the weighting factor $Weig_t(n)$) in the frequency range of 2.4 to 3.4 KHz is below 1.75dB. The corresponding values for "high quality" LPC quantisation are 10dB≤LogSegSNR≤9.5dB and 2dB≤AverSDM≤1.75dB.

4. COMPUTER SIMULATION RESULTS

The proposed SMQ approach has been simulated for different values of K, m(k) and N. The corresponding SMQ codebooks have been designed using, for training, 150 min duration of multi-speaker, multi-language speech material. In addition, several minutes of "out of training" speech from two male and two female speakers was used to evaluate the performance of various SMQ configurations. Also a conventional 3-way {3,3,4} Split-VQ scheme has been employed as a benchmark in these experiments.

The limitation of SDM to adequately reflect subjective performance became clear at the early stages of experimentation. For example a 3-way Split-VQ scheme operating at 22 bits/frame provided the same AverSDM value of 1.67dB with that obtained from a 18bits/frame Single Track (K=10, N=4) SMQ quantiser (ST-SMQ, N=4). Subjectively however, ST-SMQ, N=4 produced considerably better speech quality.

The crucial role of the weighting functions used in Equation 3 is highlighted in Figure 3, where LogSegSNR values are plotted using different numbers of bits/frame for ST-SMQ, N=4 with or without weighting in the distortion measure. The 0.65dB difference in the two curves corresponds to a net gain of 2bits/frame.

Figure 4 illustrates the LogSegSNR performance of several systems, as a function of bits/frame. An increase of N from 3 to 4 provides a 2bits/frame advantage whereas a further increase to N=5 provides a smaller gain of 0.5bits/frame. Thus with N=4 and a basic LPC frame of 20 msec duration, the system operates effectively at a rate of 12.5 segments/sec. This is comparable to the average phoneme rate and seems to be the segment length that exploits most of the existing interframe LPC correlation. Results are also included in Figure 4 for Double Track SMQ (DT-SMQ) systems. These offer improved performance, as compared to ST-SMQ schemes. DT-SMQ quantisers can deliver an advantage of 12bits/frame as compared to conventional Split-VQ.

Figure 5 illustrates storage requirements in terms of number of codebook elements for different SMQ configurations.

5. CONCLUSIONS

In this paper we have presented a new and efficient low bit rate LPC quantisation method called Split Matrix Quantisation. The direct LSP representation has been found to be particularly useful to SMQ. Within this framework the time/spectral domain weighting functions used in forming the required distortion measure play an essential role in the codebook design and search processes. Furthermore, a new and perceptually meaningful objective distortion metric is also proposed. "Transparent" LPC quantisation is easily obtained at 18bits/frame, whereas "high quality" LPC quantisation is achieved at bit rates in the range of 15 to 18bits/frame, using configurations with different complexity characteristics. Increased storage requirements can be reduced by quantising more coarsely unvoiced frames and by finding efficient ways to represent codevector elements. These issues are currently under investigation.

References

[1] I.A. Gerson and M.A. Jasiuk, "Vector Sum Excitation Linear Prediction (VSELP) Speech Coding at 8Kbps", Proc. ICASSP-90, pp. 461-464, 1990

[2] F. Soong and B.-H. Juang, " Optimal Quantization of LSP Parameters Using Delayed Decisions", Proc. ICASSP-90, pp. 185-188,1990

[3] K.K. Paliwal and B.S. Atal, " Efficient Vector Quantization of LPC Parameters at 24 bits/frame", Proc. ICASSP-91, pp. 661-664,1991

[4] P. Hedelin, " Single Stage Spectral Quantisation at 20 bits", Proc. ICASSP-94, pp. I.525-I.528,1994

[5] C.S. Xydeas and K.K.M. So, "A Long History Quantisation Approach to Scalar and Vector Quantisation of LSP Coefficients", Proc. ICASSP-93 , pp. II.1-II.4, 1993

[6] M. Yong, G. Davidson and A. Gersho, "Encoding of LPC Spectral Parameters Using Switched-Adaptive Interframe Vector Prediction", Proc. ICASSP-88, pp.402-405, 1988

[7] R.A. Salami, L. Hanzo and D.G. Appleby, "A computational Efficient CELP Coder with Stochastic Vector Quantisation of LPC parameters", 1989 URSI Issue, pp. 140-143, 1989.

[8] C. Tsao and R.M. Gray, "Matrix Quantizer Design for LPC Speech Using the Generalised Lloyd Algorithm", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-33, 1985.

[9] G.S. Kang and L.J. Fransen, "800-B/S Voice Encoding Algorithm", Proceedings of the Tactical Communications Conference. Tactical Communications: Technology in Transition., Vol. 1 pp. 57-64, 1992.

[10] M. Honda and Y. Shiraki, "Very Low-Bit-Rate Speech Coding", Advances in Speech Signal Processing, Editor S. Furui, M.M. Sondhi, Mariel Dekker Inc., 1991.

[11] G.S. Kang and L.J. Fransen, "Application of Line Spectrum Pairs to Low-Bit-Rate Speech Encoders", Proc. ICASSP-85, pp. 7.3.1-7.3.4, 1985.

[12] H.J. Coetzee and T.P. Barnwell III, "An LSP Based Speech Quality Measure", Proc. ICASSP-89, pp. 596-599, 1989.

[13] T.G. Champion, R.J. McAulay T.F. Quatieri, "High-Order All-Pole Modelling of the Spectral Envelope", Proc. ICASSP-94, Proc. ICASSP-89, Vol. I, pp. 529-532,1994

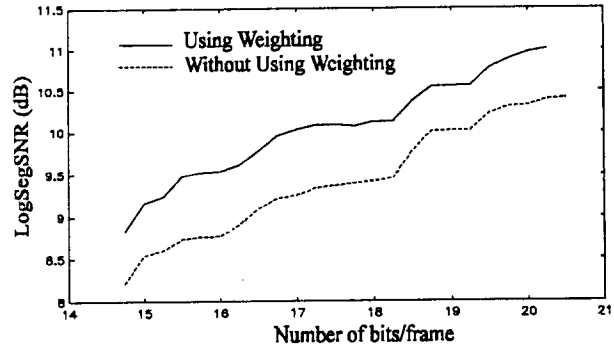

Figure 3: LogSegSNR vs. bit rate for ST-SMQ (N=4) with and without weighting in the distortion measure.

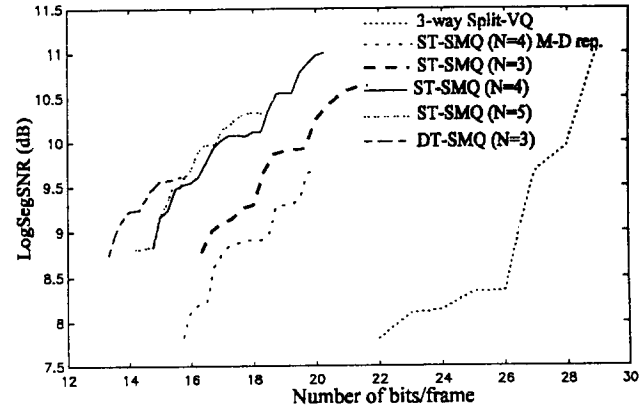

Figure 4: LogSegSNR vs. bit rate for ST-SMQ (n=3,4,5), ST-SMQ (N=4) Using Mean-Difference representation (M-D rep.), DT-SMQ (N=3) and 3way Split-VQ.

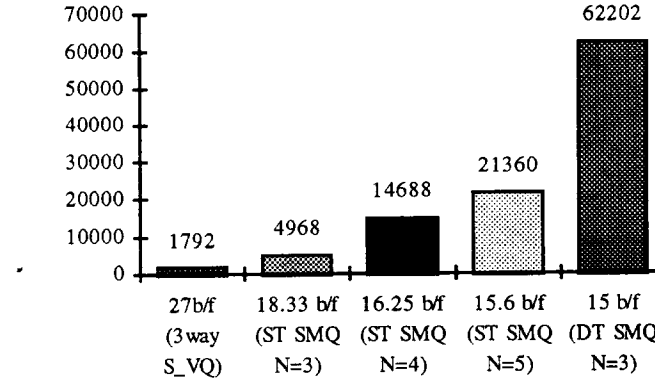

Figure 5 : Storage requirement for "high quality" LPC Quantisation