COMPARATIVE STUDY OF WIDEBAND SPEECH SPECTRAL QUANTISATION SCHEMES

G. Guibé, H.T. How, and L. Hanzo

Dept. of Electr. and Comp. Sc.,Univ. of Southampton, SO17 1BJ, UK.
Tel: +44-1703-593 125, Fax: +44-1703-594 508
Email: lh@ecs.soton.ac.uk http://www-mobile.ecs.soton.ac.uk

ABSTRACT

In this treatise a range of *Line Spectrum Frequency* (LSF) Vector Quantization (VQ) schemes were studied comparatively, which were designed for *wideband speech codecs*. Both predictive arrangements and memoryless schemes were investigated. Specifically, both memoryless Split Vector Quantization (SVQ) and Classified Vector Quantization (CVQ) were studied. These techniques exhibit a low complexity and high channel error resilience, but require high bit rates for maintaining high speech quality. By contrast, Predictive Vector Quantizers (PVQ) offer a reduced Spectral Distortion (SD) performance, although they are sensitive to channel error propagation. It is shown that the family of the so-called Safety-Net Vector Quantizers (SNVQ) offers a good design compromise in this sense.

1. INTRODUCTION

In wideband speech codecs a high number of spectral coefficients - typically - 16 has to be quantized, in order to represent the spectrum up to frequencies of 7kHz. However, the Line Spectral Frequency (LSF) coefficients above 4kHz are less amenable to Vector Quantization (VQ) than their low-frequency counterparts. Table 1 summarizes most of the recent approaches to wideband speech spectral quantization found in the literature. In this contribution we provide a comparative study of a range of efficient wideband LSF vector quantisers, documenting their design trade-offs and performance.

2. STATISTICAL PROPERTIES OF WIDEBAND LSFS

The employment of the LSF [7, 8, 9, 10, 11] representation for quantization of the LPC parameters is motivated by their statistical properties. Figure 1 shows the Probability Density Functions (PDFs) of 16 wideband speech LSFs over the interval of 0-8kHz. These different PDFs have to be taken into account in the design of the associated LSF quantizers.

	Quantization Scheme	No. of Bits per Frame
Harborg et al [1]	scalar	60, 70 and 80
Lefebvre et al [3]	Split VQ	49
Paulus et al [4]	Predictive VQ	44
Chen et al [2]	Split-VQ	49
Ubale et al [6]	Multi-stage VQ	28
Combescure et al [5]	Multi-stage Split VQ	33 at 16kbit/s 43 at 24 kbit/s

Table 1: Overview of Wideband LPC Quantizers.

The essential motivation of vector quantization is the exploitation of the relationship between the LSFs in both the frequency and the time domain. Figure 2 shows the time-domain evolution of the wideband (WB) speech LSF traces, demonstrating their strong correlation in consecutive frames in the time-domain, which is often referred to as their inter-frame correlation. Similarly, it demonstrates within each speech frame the ordering property of neighbouring LSF values, which is also referred to as intra-frame correlation, motivating the employment of vector quantization.

3. WIDEBAND LSF VECTOR QUANTIZERS

3.1. Memoryless Vector Quantization

The so-called Nearest Neighbour Vector Quantization (NNVQ) scheme [14] theoretically constitutes the optimal memoryless solution for VQ. However, the high number of LSFs - typically 16 - required for wideband speech spectral quantization results in a complexity that is irrealistic for a real time implementation, unless the 16-component LSF vector is split into subvectors. As an extreme alternative, low complexity scalar quantization constitutes the ultimate

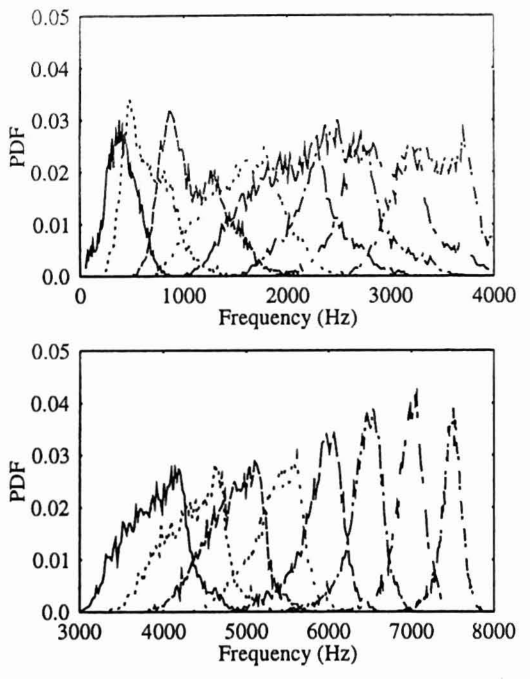

Figure 1: PDFs of the LSFs using LPC analysis with a filter order of 16, demonstrating the ordering property of the LSFs.

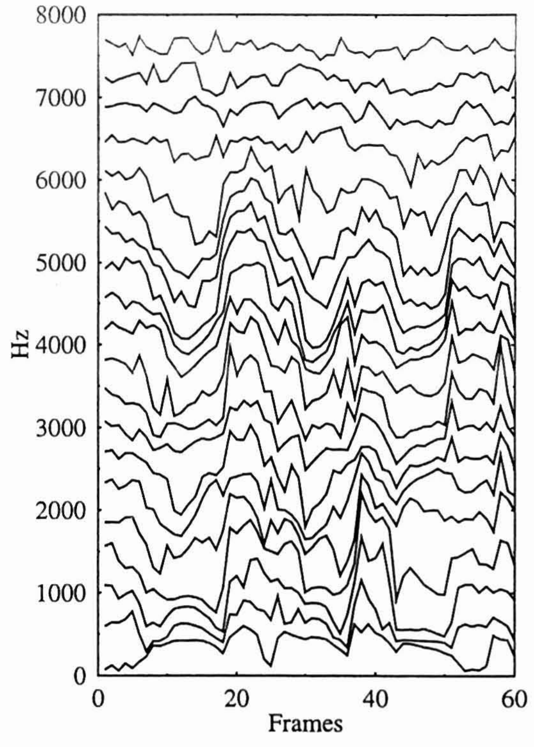

Figure 2: Traces of 16 wideband LSFs, demonstrating their inter- and intra-frame correlations.

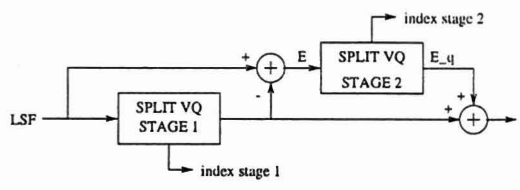

Figure 3: Schematic of the multi-stage split VQ

splitting of the original LSF vector into one-dimensional sub-vectors. This method exhibits a low complexity and a good Spectral Distortion (SD) performance [13] can be achieved using 16-entry or 4-bit codebooks for each individual LSF. Nevertheless, the large number of LSFs required in wideband speech codecs implies a requirement of 4·16=64 or 5·16=80 bits per 10 ms speech frame. As a result, the contribution of the scalar quantized LSFs to the codec's bit rate is 6.4 or 8 kbit/s. Slight improvements can be achieved using a non-uniform bit allocation, when more bits are allocated to the perceptually most significant LSFs.

Between the above extreme cases, split VQ (SVQ) aims to define a split configuration that minimizes the average SD within a given total complexity. Specifically, split vector quantization operates on sub-vectors of dimensions that can be vector quantized within the given constraints of complexity, following the schematic of Figure 3.

One of the main issues in split LSF VQ is defining the best possible partitioning of the initial LSF vector into sub-vectors. Since the high frequency LSFs typically exhibit a different statistical behaviour from their low frequency counterparts, they have to be encoded separately. For linear predictive filters of order 16 the 3 highest order LSFs behave differently from the other LSFs, as exemplified by Figure 2. Hence, this leads naturally to a (13,3)-split VQ scheme. In our further discussions we will characterise the performance of various LSF VQs in terms of the PDF of their SD following the approach of Paliwal and Atal [13].

Apart from a low average SD, the various designs have to minimise the probability of high-SD speech frames associated with the so-called outliers of the PDF.

In this context Figure 4 shows the PDF of the SD using a (6,7,3)-split LSF VQ scheme, where the lower frequency 13-component sub-vector is split into two further 6- and 7-component sub-vectors, in order to reduce the implementational complexity. Seven bits, i.e 128 codebook entries were used for each sub-vector. Additionally, a (4,4,4,4)-split second stage VQ was applied according to Figure 3 using five bits, i.e. 32 codebook entries for each sub-vector. We refer to this scheme as the $[(6,7,3)_{777};(4,4,4,4)_{5555}]$ 41-bit regime. Having considered memoryless VQs, let us now focus our attention on predictive VQ schemes.

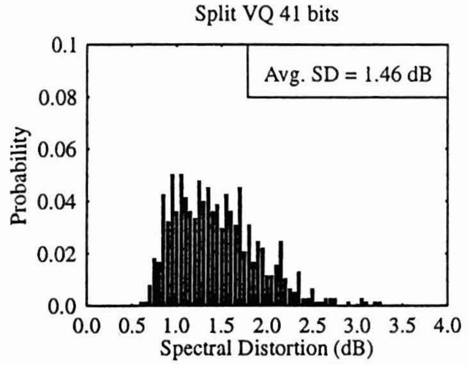

Figure 4: PDF of the SD for the 41-bit split VQ scheme using the $[(6,7,3)_{777}; (4,4,4,4)_{5555}]$ two-stage regime (Compare to Figure 6 and 9).

3.2. Predictive Vector Quantization

In this section our work evolves from memoryless vector quantization to more efficient vector quantization schemes exploiting the time-domain inter-frame correlation of LSFs. According to this approach we typically quantize a sequence of vectors, where successive vectors may be statistically dependent.

Predictive vector quantization (PVQ) constitutes a vector-based extension of traditional scalar predictive quantization. Its schematic is shown in Figure 5. PVQ schemes aim to exploit the correlation between the current vector and its past values, in order to reduce the variation range of the signal to be quantized. Provided that there is sufficient correlation between consecutive vectors and the predictor is efficient, the vector components to be quantized are expected to be unpredictable, random noise-like signals, exhibiting a reduced dynamic range. Hence, for a given number of codebook entries, PVQ is expected to give a lower SD, than non-predictive VQ.

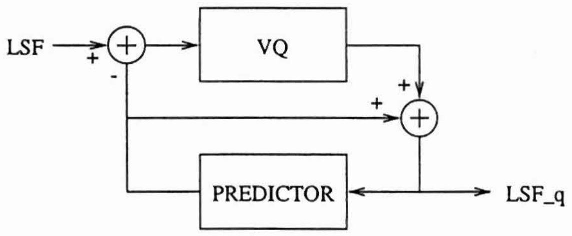

Figure 5: Schematic of a Predictive Vector Quantizer (PVQ).

Auto-Regressive (AR) predictors use recursive reconstruction of the LSFs, hence they potentially suffer from severe propagation of channel errors over consecutive frames. By contrast, a Moving-Average (MA) predictor can typically limit the error propagation to a lower number of frames, given by the predictor order. Here, however we will restrict our experiments to first order AR vector predictors.

Predictive vector quantization does not necessarily preserve the LSF's ordering property, which may result in instability of the synthesis filter, deteriorating the perceptual speech quality. In order to counteract this problem, an LSF rearrangement procedure [15] can be introduced, ensuring a minimum distance of 50Hz between neighbouring LSFs in the frequency domain.

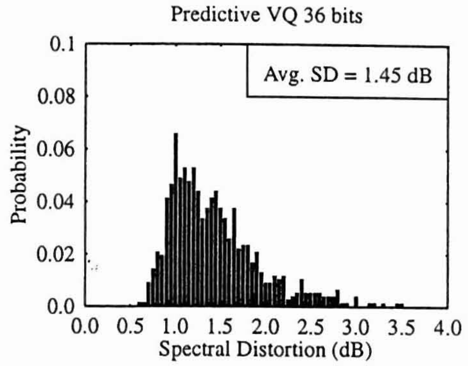

Figure 6: PDF of the SD for the 36-bit PVQ scheme (Compare to Figure 4 and 9).

Figure 6 shows the PDF of the SD using $(4,4,4,4)_{9999}$ 36-bit split vector quantization of the prediction error, employing a 9-bit codebook per 4-LSF sub-vector. This quantizer hence requires a total of $4 \cdot 9 = 36$ bits per LSF vector. Based on the above experience we conclude that our 36-bit Predictive VQ provides a gain of 5 bits per LSF vector in comparison to our previous 41-bit memoryless SVQ having a similar complexity. Equivalently, predictive VQ generates an average SD gain of approximatively 0.3 dB for a given bit rate. A deficiency of this method is its higher sensitivity to channel error propagation, although this problem can be mitigated by using MA prediction instead of AR prediction. During our investigations we noted that this scheme was sensitive to unpredictable LSF vectors generated by the rapid speech spectral changes, which increase both the average SD as well as the number of SD outliers beyond SD=2dB. This problem is addressed in the next section.

3.3. Multimode Vector Quantization

Let us now consider a range of multimode VQ schemes. When we observe these voiced/unvoiced speech transitions in the time domain, they result in the rapid changes of the LSF traces seen in Figure 2 for example around frame 20. Several methods exist for differentiating between these modes. Switched prediction is widely employed [11, 15]. In this section, we will investigate the separate encoding of the unpredictable frames due to rapid spectral changes and that of the highly-correlated frames. This can be achieved by the combination of a predictive VQ and a fixed memoryless SVQ, referred to as the so-called Safety-Net VQ (SNVQ) scheme [16, 17, 18]. In this context, we invoke a full search using both the predictive VQ and the fixed memoryless SVQ schemes for every speech frame, and the better candidate with respect to a mean-squared distortion criterion is chosen.

The safety-net VQ improves the overall robustness against outliers, which are typically due to input LSF vectors having a low correlation with the previous LSF vectors. In addition, the safety net VQ allows the PVQ to concentrate on the predictable, highly correlated frames. Hence, the variance of the LSF prediction error is reduced and a higher resolution LSF prediction error codebook can be designed. The advantage of this method is that when the inter-frame correlation cannot be successfully exploited in a PVQ scheme, the intra-frame correlation is capitalised on instead.

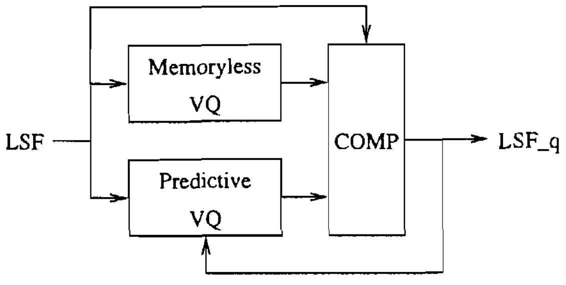

Figure 7: Schematic of the Safety Net Vector Quantizer (SNVQ) constituted by a memoryless- and a predictive-VQ.

Figure 7 shows the structure of the SNVQ scheme. Again, the input LSF vector is quantized using both predictive- and memoryless quantizers, then both quantized vectors are compared to the input vector, in order to select the better quantization scheme. The codebook index selected is transmitted to the decoder, along with a signalling bit that indicates the selected mode. The specific transmitted quantized vector is finally used by the PVQ, in order to predict the LSF vector of the next frame.

The performance difference between the memoryless SVQ and predictive VQ sections of the SNVQ suggests the employment of variable bit rate schemes, where the lower performance of the memoryless SVQ can be compensated by using a larger codebook. In our experiments below - as before - a 41-bit memoryless SVQ codebook was used. Hence, the SNVQ is characterized by its average bit rate, depending on the proportion of vectors quantized by the predictive and memoryless VQ, respectively. Eriksson, Linden and Skoglung [17] argued that the optimum performance is attained, when 50 to 75% of frames invoke the PVQ.

Figure 8 shows the proportion of frames quantized using the 28-, 32- and 36-bit PVQs in the context of SNVQ schemes employing 36-, 41- and 43-bit SVQs. We observe in Figure 8 that for a PVQ codebook size of 28 and 32 bits a relatively low proportion of the LSF vectors was quantized using the PVQ and this indicated that its codebook size was too small, failing to outperform the memoryless 36-, 41- or 43-bit SVQs. Accordingly, only the 36-bit PVQ was

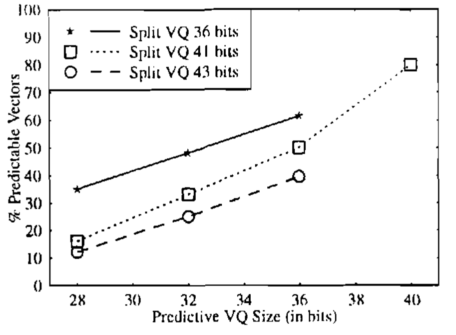

Figure 8: Proportion of frames using PVQ in various SNVQ schemes, employing memoryless SVQs of 36, 41 and 43 bits.

deemed suitable. This figure illustrates that if the predictive VQ exhibits a low performance compared to the memoryless SVQ, i.e. the proportion of its utilization tends to zero, the SNVQ will tend to behave like a simple memoryless SVQ. Alternatively, if the memoryless SVQ exhibits a low performance compared to the PVQ, i.e. the proportion of PVQ LSF vectors tends to 100%, the SNVQ will tend to behave like a PVQ.

The individual PVQ and memoryless SVQ schemes employed so far were designed independently from each other, hence the resulting scheme is sub-optimal. Furthermore, both quantizers were designed without distinction between predictable and unpredictable LSF vectors. Hence, their optimization will aim, on one hand, to have the PVQ focusing on predictable frames, which generate LSF prediction errors with a low variation range. On the other hand, the memoryless SVQ codebook is to be matched to the distribution of the unpredictable LSF vectors in the p-dimensional LSF space. In order to obtain an optimal SNVQ we will proceed as follows:

1) The original training sequence T is passed through our previously used individual sub-optimum codebook based SNVQ, in order to generate the sub-training sequences T_{PVQ} and T_{SN} of vectors, quantized using either the predictive VQ or the memoryless SVQ, respectively, depending on which generated a lower SD.

2) Then codebooks for both the PVQ and the memoryless SVQ are designed using the sub-training sequences generated above.

Our results to be highlighted with reference to Table 2 show that the optimized PVQ results in significant improvements, but only a modest further gain was obtained with the aid of the Safety-Net approach, invoking the optimised memoryless SVQ. Clearly, optimization is the main issue in SNVQ design, requiring the joint design of both parts of the SNVQ. We designed a [36,36]-bit and a [36,41]-bit scheme, where the first bracketed number indicates the number of bits assigned to the PVQ, while the second one that of the memoryless SVQ. Again, the performance of these schemes is summarised in Table 2. In both cases a SD gain of about 0.15 dB was obtained upon the joint optimisation of the component VQs, as seen in Table 2. In addition, the number of outliers between 2 and 4 dB was substantially reduced and all the outliers over 4 dB were removed.

Scheme	Avg. SD (dB)	Outliers (%) 2-4 dB	Outliers (%) >4 dB
[36,36] SNVQ scheme			
non-optimized	1.34	7.19	0.12
optimized	1.17	2.18	0
[36,41] SNVQ scheme			
non-optimized	1.25	4.5	0.12
optimized	1.09	0.38	0

Table 2: Optimization effects for the [36,36] and [36,41] SNVQ schemes.

We found that the optimization slightly increased the proportion of frames quantized using the PVQ. For our [36,36] SNVQ scheme, this proportion increased from 67% to 74%. Similarly, for the [36,41] SNVQ scheme constituted by the 36-bit PVQ and 41-bit memoryless SVQ respectively, this proportion increased from 50% to 60%. Hence, in case of such switched variable bit rate schemes, the optimization tends to reduce the average SNVQ bit rate, since the PVQ requires less bits, than the memoryless SVQ.

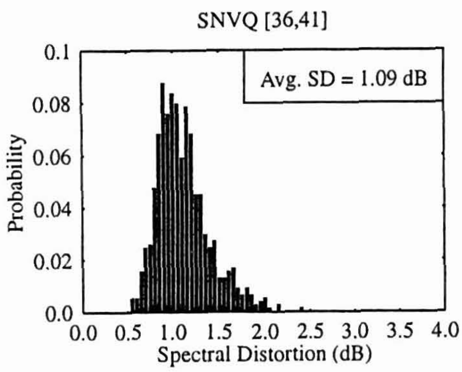

Figure 9: PDF of the SD for the [36,41] bit SNVQ scheme (Compare to Figure 4 and 6).

Figure 9 shows the PDF of the SD for the [36,41] SNVQ scheme, indicating a significant SD PDF enhancement compared to both the memoryless SVQ and the PVQ. In addition, this system improves the robustness against channel errors, since the propagation of bit errors was limited due to the low number of consecutive employment of the PVQ. Clearly, the SNVQ enabled an efficient exploitation of both the inter-frame correlation and the intra-frame correlation of LSF vectors. Its main deficiency is the increased complexity of the codebook search procedure, requiring twice as many comparisons as the memoryless SVQ or the PVQ.

4. SIMULATION RESULTS AND SUBJECTIVE EVALUATIONS

Figure 10 summarizes the performance of the split memoryless SVQ, the PVQ and the SNVQ. As observed in the figure, the SD results for the memoryless SVQ are more modest and in general a better performance was obtained by using the predictive quantization schemes. This figure illustrates a difference of 4 or 5 bits between the memoryless SVQ and the PVQ for the same SD. The three SD curves corresponding to the SNVQ schemes using 28-, 32- and 36-bit PVQs in conjunction with various associated memoryless SVQ configurations are also shown in Figure 10. For the SNVQ using 28- and 32-bit PVQs, the lines crossing the PVQ performance curve drawn using a solid line indicate that at this stage the PVQ starts to attain a better performance, than the SNVQ for the equivalent bit rate. Hence, in this scenario there is no benefit from employing SNVQ schemes using 28- and 32-bit PVQs beyond this cross-over point. A consistent SD gain in comparison to the PVQ is only ensured for the SNVQ using the 36-bit PVQ. In this case a 2-bit reduction in the number of required coding bits was obtained. Informal listening tests have shown that the best perceptual performance was obtained by employing the [36,41] SNVQ scheme.

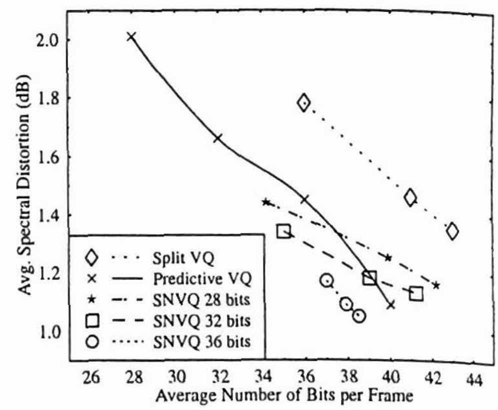

Figure 10: Average SD of the various vector quantizers considered in this study

Scheme	No. of Bits	Avg. SD (dB)	Outliers (%) 2-4 dB	Outliers (%) >4 dB
PVQ	40	1.09	4.24	0
SNVQ	38	1.09	0.38	0

Table 3: Transparent quantization schemes.

Table 3 details the characteristics of two high-quality quantization schemes. The first configuration utilised a $(4,4,4,4)_{10,10,10,10}$ PVQ scheme employing $4 \cdot 10 = 40$ bits and the second scheme used a [36,41] SNVQ arrangement with an average of 38 bits. Although both schemes have a similar average SD, the SNVQ provides a large reduction in the number of SD outliers between 2 and 4 dB, which have a significant effect on the perceptual speech quality. A high speech quality was also obtained for the [36,36] fixed bit rate SNVQ, as shown in Table 2.

5. CONCLUSIONS

In this contribution, we have comparatively studied various predictive and memoryless vector quantizers. In the context of memoryless VQ, a $[(6,7,3)_{777}; (4,4,4,4)_{5555}]$ 41-bit multi-stage split vector quantizer was designed. This method enabled a simple implementation. In the context of 41-bit predictive vector quantization a SD quality enhancement was achieved compared to memoryless schemes, or alternatively the number of bits could be reduced to 36, while maintaining a similar average SD. The associated SD PDFs were portrayed in Figures 4, 6 and 9, while their salient features were summarised in Tables 2 and 3. Unfortunately, the channel error sensitivity increased due to potential error propagation. Lastly, we combined both the memoryless- and the predictive approaches in a SNVQ scheme. Even though the SNVQ scheme increased the complexity, it improved significantly the SD performance and mitigated the propagation of channel errors. Our future work considers the design trade-offs of wideband backwards adaptive speech codecs and transform codecs.

6. ACKNOWLEDGEMENTS

The financial support of the following organisations is gratefully acknowledged: Motorola ECID, Swindon, UK; European Community, Brussels, Belgium; Engineering and Physical Sciences Research Council, Swindon, UK; Mobile Virtual Centre of Excellence, UK.

7. REFERENCES

[1] E. Harborg, J.E. Knudsen, A. Fuldseth, and F. Johansen, "A Real-Time Wideband CELP Coder for a Videophone Application", in *Proc. ICASSP*, pp.121-124, 1994.

[2] J-H. Chen and D. Wang, "Transform Predictive Coding of Wideband Speech Signals", in *Proc. ICASSP*, pp. 275- 278, 1996.

[3] R. Lefebvre, R. Salami, C. Laflamme , and J-P. Adoul, "High Quality Coding of Wideband Audio Signals Using Transform Coded Excitation (TCX)", in *Proc. ICASSP*, pp. 193-196, 1994.

[4] J. Paulus and Schnitzler, "16 kbit/s Wideband Speech Coding Based on Unequal Subbands", in *Proc. ICASSP*, pp. 255-258, 1996.

[5] P. Combescure, J. Schnitzler, K. Fischer, R. Kirchherr, C. Lamblin, A. Le Guyader, D. Massaloux, C. Quinquis, J. Stegmann, and P. Vary, "A 16, 24, 32 Kbit/s Wideband Speech Codec Based on ATCELP", in *Proc. ICASSP*, Phenix, 1999.

[6] A. Ubale and A. Gersho, "A Multi-Band CELP Wideband Speech Coder", in *Proc. ICASSP*,pp. 1367-1370, 1997.

[7] F. Itakura, "Line Spectrum Representation of Linear Predictor Coefficients of Speech Signals", in *J. Acoust. Soc. Amer.*, Vol. 57, pp. S35(A), 1975

[8] L. Rabiner, M. Sondhi and S. Levinson, "Note on the Properties of a Vector Quantizer for LPC Coefficients", in *The Bell System Technical Journal*, Vol. 62, No. 8, pp. 2603-2616, October 1983.

[9] F. Soong and B.-H Juang, "Line Spectrum Pair (LSP) and Speech Data Compression",in *Proc. ICASSP*, pp. 1.10.1-1.10.4, San Diego, 1984.

[10] R. Steele and L. Hanzo, *Mobile Radio Communications*, IEEE Press, 1999, 2nd Edition.

[11] A. Kondoz, *Digital Speech, Coding for Low Bit Rate Communication System*, John Wiley, 1994.

[12] "7 kHz Audio Coding within 64 kbit/s", CCITT Recommendation G.722, 1988.

[13] K. Paliwal and S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame", in *IEEE Trans. Speech and Audio Processing*, Vol. 1, January 1993.

[14] A. Gersho and M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1991.

[15] ITU-T, "Recommendation G.729-Coding of speech at 8 kb/s using conjugate algebraic-code-excited linear-prediction (CS-ACELP)", March 1996.

[16] T. Eriksson, J. Linden, and J. Skoglung, "A Safety-Net Approach for improved exploitation of speech correlation", in *Proc. International Conference on Digital Signal Processing*, Vol. 1, pp. 96-101, 1995.

[17] T. Eriksson, J. Linden, and J. Skoglung, "Exploiting Interframe Correlation in Spectral Quantization. A Study of Different Memory VQ Schemes", in *Proc. ICASSP*, pp. 765-768, May 1996.

[18] H. Zarrinkoub and P. Mermelstein, "Switched Prediction and Quantization of LSP Frequencies", in *Proc. ICASSP*, pp. 757-764, May 1996.