Wideband speech coding, speech spectral quantisation

Speech Spectral Quantizers for Wideband Speech Coding

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

November 15, 2000

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 an enhanced Spectral Distortion (SD) performance, although they are sensitive to channel error propagation. It is shown that the family of so-called Safety-Net Vector Quantization (SNVQ) schemes offers a good design compromise, providing an extension to memory-based PVQ, and thereby improving the performances both over noisy and noiseless channels.

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. The approach employed by Harborg textitet al. [1] is based on Scalar Quantization (SQ). However, the resulting bit rate is excessive, requiring 3 or 4 bits for each LSF. Chen textitet al. [2] as well as Lefebvre textitet al. [3] utilized low dimensional split VQ. For instance, a $(2,2,2,2,2,3,3)_{7777777}$ split VQ is invoked in their approach, where only 2 or 3-dimensional VQs are used, employing 7 bits - i.e. 128 codebook entries - per sub-vector. This reduces the number of bits allocated to the LSF quantization compared to SQ, although the resulting number of bits still remains somewhat high, namely $7 \cdot 7$=49. Clearly, these approaches are simple, but a large number of bits is required.

Paulus textitet al. [4] proposed a coding scheme based on sub-band analysis of the speech signal. The speech signal was split into two unequal sub-bands, namely 0-6kHz and 6-7kHz. LPC analysis was only invoked in the lower band, using 14 LSF coefficients quantized with 44 bits per 15 ms. The quantization scheme employed inter-frame moving average prediction and split vector quantization. In the 6-7kHz higher sub-band only the signal energy was encoded using 12 additional bits. Following a similar ap-

	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.

proach Combescure textitet al. in [5] described a system based on two sub-bands, where the lower band (0-5kHz) applied a 12-th order LP filter with its coefficients quantized using 33 bits. The upper band (5-7kHz) uses an 8-th order LP filter encoded with 10 bits, but these coefficients were only transmitted in the higher bit rate mode of the coder, namely at 24 kbit/s. The lower-band coefficients were quantized using Predictive Multi-Stage Split Vector Quantization (MSVQ). These types of LSF quantizers are not directly amenable to employment in fullband wideband speech codecs. However, the approach using separate coding of the higher- and lower-band LSFs can be helpful in general for LPC quantization.

Finally, Ubale textitet al. in [6] proposed a scheme using predictive MSVQ of seven stages employing four bits each. This method employed a so-called multiple survivor method, where four - rather than one - residual survivors were retained at each pattern-matching stage and were then tested at the next pattern-matching stage. The final decision was taken at the last VQ stage as to which of the split vector combinations gave the lowest quantization error. In addition, the MSVQ was designed by a joint optimization procedure, clearly demonstrating the advantages of using schemes, which predictively exploit the knowledge of the signal's past history, in order to improve the coding efficiency.

Having reviewed the background of wideband speech spectral quantization, we now focus our attention on the statistical properties of the wideband speech LSFs, which render it attractive for vector quantization.

1.1 Statistical Properties of Wideband LSFs

The employment of the LSF [7, 8, 9] 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. Their different PDFs have to be taken into account in the design of the 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.

Intra-frame correlation motivates the employment of vector quantization, since it enables a mapping that matches the multi-dimensional LSF distribution. We observe at the top of Figure 2 that the correlation of the individual LSFs within a given speech frame tends to decrease, as the fre-

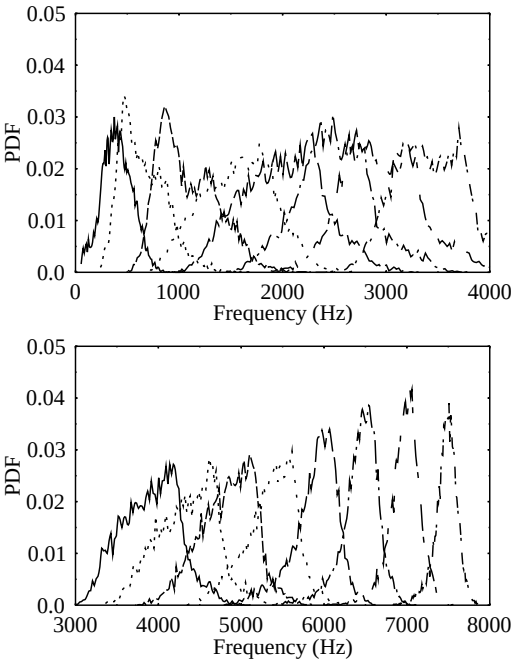


Figure 1: PDFs of the LSFs using LPC analysis with a filter order of 16, demonstrating the ordering property of the LSFs.

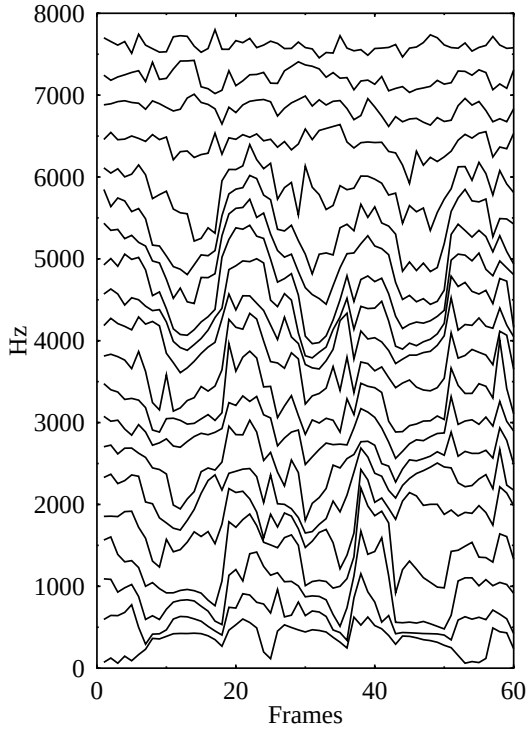


Figure 2: Traces of 16 wideband LSFs, demonstrating their inter- and intra-frame correlations.

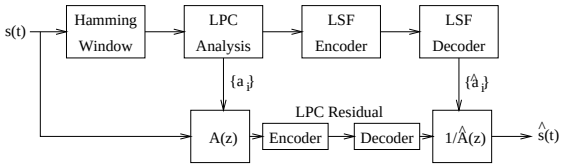


Figure 3: Evaluation of the perceptual speech quality after applying LSF vector quantization

quency increases, i.e. higher frequency LSFs are more statistically independent of each other, although they still obey the ordering property. This clearly manifests itself for example around frame 18 in Figure 2. The highest frequency LSFs describe the noisy high frequency bands of the speech signal, which typically appear to be noise-like. This characteristic will mostly be exploited in the design of memoryless VQ schemes.

Inter-frame correlation of the LSFs can be exploited by interframe predictive vector quantization schemes having memory, where predictions of the current LSF values are employed, in order to reduce the variance of the vector we want to quantize. Finally, when rapid spectral changes are observed in the LSF traces, affecting both their intra- and inter-frame correlation, various multi-mode schemes can be invoked, as we will show during our further discourse.

1.2 SPEECH CODEC SPECIFICATIONS

The design of speech codecs is based in general on a trade-off between the conflicting factors of perceptual speech quality, the required bit rate, the channel error resilience and the implementational complexity. Wideband speech coding [12] aims to provide a better perceptual quality, than narrowband speech codecs. Hence, a fine quantization of the LPC parameters is required.

Listening tests using the scheme depicted in Figure 3 indicate that the transparency criterion formulated by Paliwal and Atal [13] in the context of narrowband speech codecs is also relevant in wideband scenarios. This criterion uses a Spectral Distortion (SD) measure given by

$$SD^2 = \frac{1}{f_s}\int_0^{f_s}\left[10log_{10}(P(f)) - 10log_{10}(\widehat{P}(f))\right]^2 df,$$

where $P(f)$ and $\widehat{P}(f)$ are the amplitude spectra of the original and reconstructed signal, respectively. The required criteria are satisfied, if an average SD of about 1 dB is maintained and there are only a few 'outliers' between SD=2 and 4 dB, while there are no outliers in excess of SD=4 dB. In addition, an important issue in speech quality terms is the preservation of the stability of the Short Term Predictor (STP). The STP filter's stability has a dramatic influence on the reconstructed speech quality, which is guaranteed by preserving the ordering property of the LSFs.

Every codec designed for transmission over noisy channels has to exhibit a good robustness against channel errors. The effect of transmission errors is characterized by their immediate impact on both the present speech frame and also on the forthcoming frames. The effect of channel errors and the associated bit sensitivity issues will be discussed in Section 4. Complexity reduction is also of high importance for real time applications. The codebook storage requirements and codebook search complexity are the main factors to be taken in consideration in the field of vector quantization.

In our investigations, training and testing of the vector quantization schemes was performed using wideband speech files from the TIMIT database [15]. An LPC analysis of order $p = 16$ was performed every 10 ms, using a 15 ms Hamming analysis window. A 30 Hz bandwidth expansion was applied to each pole of the LPC coefficient vector. A training set of 14000 LSF vectors was generated from speech files recorded from American male and female speakers. However, for the sake of reduced computing time, our codebook computations using the Generalized Max Lloyd algorithm [14], were usually processed using 7000 LSF vectors. In addition, a test set of 778 LSF vectors was used for the evaluation of the resulting perceptual quality.

Let us now examine in the next section a few wideband LSF vector quantization schemes.

2 WIDEBAND LSF VECTOR QUANTIZERS

2.1 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 not realistic 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 splitting of the original LSF vector into reduced-dimension sub-vectors. This method exhibits a low complexity and a good SD performance can be achieved using 16-entry or 4-bit codebooks. 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

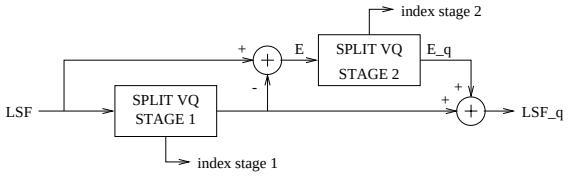


Figure 4: Schematic of the multi-stage split VQ

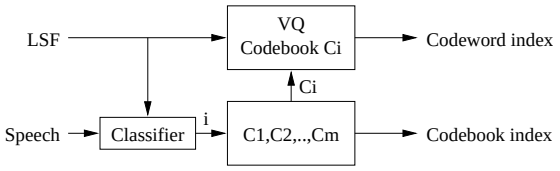


Figure 6: Schematic of the Classified VQ

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 4.

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. Figure 5 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 4 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.

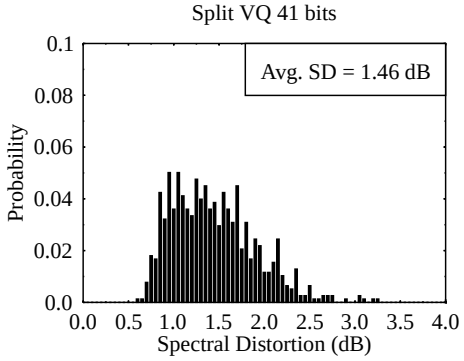


Figure 5: PDF of the SD for the 41-bit split VQ scheme using the $[(6,7,3)_{777};(4,4,4,4)_{5555}]$ two-stage regime (Compare to Figure 10 and 13).

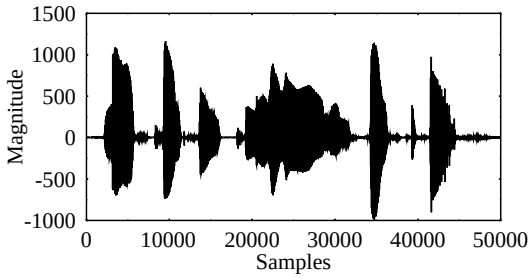
The lower intra-frame correlation of the higher frequency LSFs imposes a high bit rate requirement on the SVQ in the light of the relatively low energy contained in the corresponding speech band (typically less than 1%). Although split VQ schemes are attractive in complexity terms and can preserve the LSFs ordering property, they often fail to reach the target SD within a low bit rate budget.

The introduction of LSF Classified Vector Quantization (CVQ) [14] aims to assign the LSF vectors into classes having a particular statistical behavior, in an effort to improve the coding efficiency.

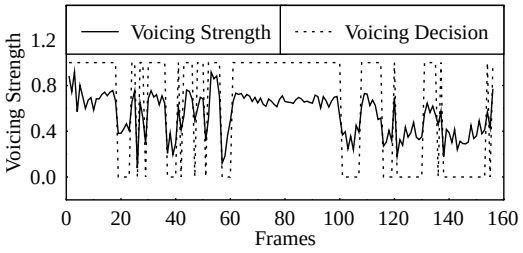
In Figure 6 the LSF vectors are classified into one of m categories $C_1 \cdots C_m$ and then a reduced-size codebook C_m, which reflects the statistical properties of class m is searched in order to find the best matching codebook entry for the unquantized LSF vector. Clearly, this scheme searches a reduced-size codebook, reducing the matching complexity and the quantization precision in comparison to a VQ using no pre-classification before quantization. In the context of wideband speech LSF quantization, we wish to find a classification of the LSFs, which can provide a more efficient representation of the vector to be quantized, than the previous SVQ. Accordingly, the main issue in classified vector quantization is the design of an accurate classifier. In this context, we briefly investigate the performance of a voiced/unvoiced classifier.

The problem of voicing detection can be solved upon invoking an auto-correlation based pitch detector [11], exploiting the waveform similarities between the original speech and its pitch-duration shifted version. The highest correlation between these two signals is registered, when their displacement corresponds to the pitch. Figure 7(a) shows a low-pass filtered speech waveform bandlimited to 900Hz, which was subjected to autocorrelation-based voicing-strength evaluation and thresholding at a normalised cross-correlation of 0.5, in order to generate the binary voiced/unvoiced (V/UV) decisions seen in Figure 7(b).

Figure 8 demonstrates the relevance of this approach, portraying - as an illustrative example - the scatter diagrams of the first two LSFs after classification. For both diagrams, the unoccupied bottom right corner region manifests the dependency between the LSFs due to their ordering property. The first two LSFs of voiced frames at the left of Figure 8 are centred around two clusters. One corresponding to the low frequency LSF 1 occurences, where LSF 2 appears near constant. The other voiced frame cluster corresponds to frames, where LSF 1 and 2 exhibit similar values, creating a near-linear cluster along the 'ordering property border'. The unvoiced frames at the right of Figure 8 appear more scattered, although they also exhibit an apparent,



(a) Low-pass filtered speech signal



(b) Voicing Strength and the associated binary Voicing Decisions

Figure 7: V/UV speech classification using low-pass filtering of the speech to 900Hz and auto-correlation based pitch detection.

but less pronounced clustering along the ordering property border.

Voiced and unvoiced LSFs do not necessarily exhibit a totally different statistical behavior in their clusters along the ordering property border in Figure 8. However, the typically more concentrated clusters of the voiced LSF frames can be typically more accurately vector quantized, whereas the somewhat more scattered occurences of the unvoiced frames' LSFs are expected to be less amenable to CVQ. Similar scatter diagrams can be obtained also for higher frequency LSFs, although the pronounced difference between voiced and unvoiced frames tends to decrease, as the frequency increases. This is directly related to the less pronounced correlation between neighbouring LSFs for the higher frequencies of the 8kHz range.

Although our simulations using this CVQ gave better SD results, than the previously discussed Split VQ, the overall scheme presents shortcomings. Specifically, if the speech frame classification is carried out before the LSF quantization, classification errors at the voiced/unvoiced speech boundaries increase the average SD, as well as the number of outliers. At the decoder, this method has to rely on the voiced/unvoiced information extracted from the excitation

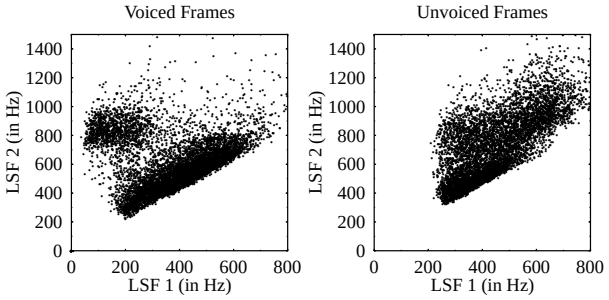


Figure 8: Scatter diagrams of the first two LSFs for WB voiced and unvoiced frames.

signal in order to reconstruct the LSF coefficients, unless the V/UV mode is explicitly signalled to the decoder. Alternatively, if the V/UV classification is processed after LSF quantization upon selecting the mode having the lower SD, no classification errors occur, although one bit per speech frame is required for transmitting the V/UV mode selection. When using the $[(6,7,3)_{777};(4,4,4,4)_{5555}]$ 41-bit Split LSF VQ for each mode, an average SD of 1.15 dB is obtained upon invoking a mode selection bit, whereas an average SD of 1.35 dB is achieved using the pitch-detection based classification.

Additionally, it is difficult to proceed to a joint optimization of both the voiced and the unvoiced codebooks, since there are regions of the LSF domain where both types of LSFs can be located. The LSF clusters, which are encountered in both modes, are quantized independently by the voiced codebook and the unvoiced codebooks. Hence the same sub-domain of the LSF space is mapped twice by the quantization cells of both modes. This leads to a sub-optimal quantization of this area. Let us now consider predictive VQ schemes.

2.2 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 9. 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, ex-

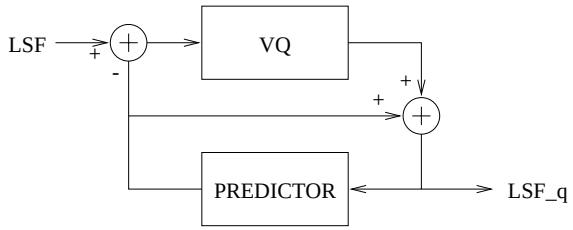


Figure 9: Schematic of a Predictive Vector Quantizer (PVQ).

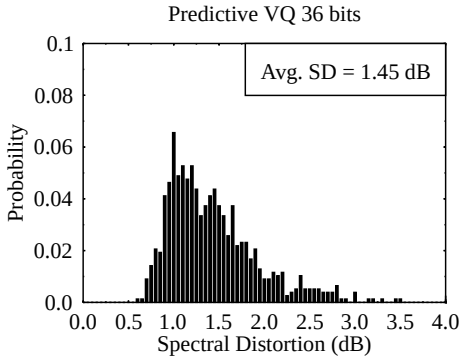


Figure 10: PDF of the SD for the 36-bit PVQ scheme (Compare to Figure 5 and 13).

hibiting a reduced dynamic range. Hence, for a given number of codebook entries, PVQ is expected to give a lower SD, than non-predictive VQ.

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. For example, a fourth-order MA predictor was succesfully employed in the G.729 narrowband standard speech codec. Other methods of limiting bit error propagation have also been studied by Eriksson *et al.* [19], in the context of narrowband speech. Here, however we will restrict our experiments to first order AR vector predictors.

Predictive vector quantization does not necessarily preserve the LSFs' ordering property. This may results in instability of the STP filter, deteriorating the perceptual quality. In order to counteract this problem, an LSF rearrangement procedure [16] can be introduced, ensuring a minimum distance of 50Hz between neighbouring LSFs in the frequency domain.

Figure 10 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 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.

2.3 MULTIMODE VECTOR QUANTIZATION

Our previous classified vector quantization scheme has primarily endeavored to define V/UV correlation modes. 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, 16]. 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 [17, 18, 19, 20]. 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. Another interesting aspect to consider is the performance of the SNVQ scheme for transmission over noisy channels, which will be discussed in Section 4. In a memory-based PVQ scheme, a bit error will lead to error propagation over consecutive frames. By invoking a memoryless SVQ scheme together with a memory-based scheme, error propagation will be cancelled every time an entry from the Memoryless SVQ codebook is selected and correctly transmitted to the decoder.

Figure 11 shows the structure of the SNVQ scheme. Again, the input LSF vector is quantized using both predictive- and memoryless quantizers, then both quantized vectors

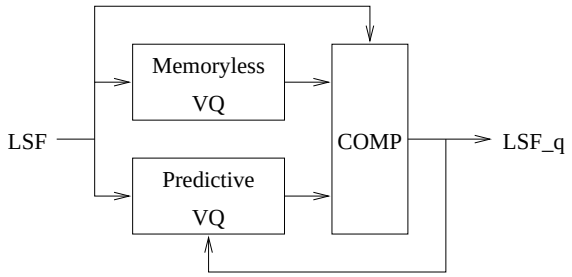


Figure 11: Schematic of the Safety-Net Vector Quantizer (SNVQ) constituted by a memoryless- and a predictive-VQ.

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 memoryless SVQ 41-bit 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 Skoglund [19] argued that the optimum performance is attained, when 50 to 75% of frames invoke the PVQ.

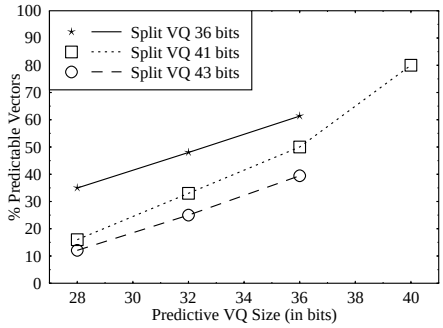


Figure 12: Proportion of frames using PVQ in various SNVQ schemes, employing memoryless SVQs of 36, 41 and 43 bits.

Figure 12 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 12 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 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.

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

Scheme	Avg. SD (dB)	Outliers (%)	
		2-4 dB	>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.

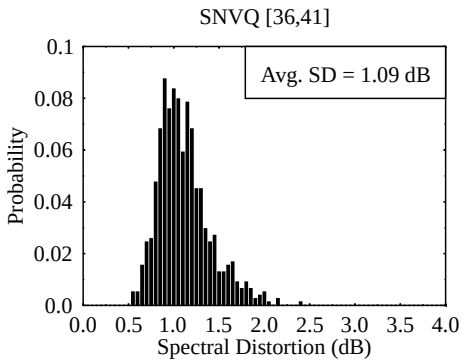


Figure 13: PDF of the SD for the [36,41] bit SNVQ scheme (Compare to Figure 5 and 10).

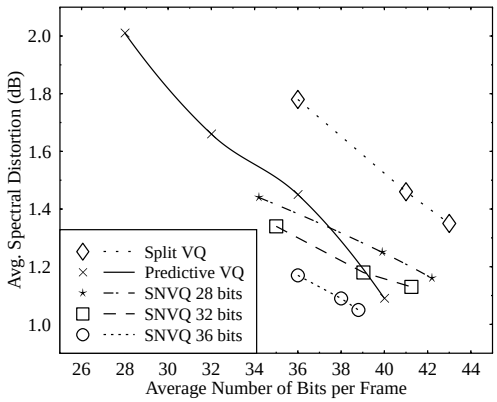


Figure 14: Average SD of the various vector quantizers considered in this study

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

Table 3: Transparent quantization schemes.

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 13 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.

3 SIMULATION RESULTS

Figure 14 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 memoryless SVQ configurations employing 36, 41 and 43 bits are also shown in Figure 14. The x-axis, which is characterised by the average number of bits per frame in Figure 14 is obtained by taking into account the proportion of utilization for the SVQ and PVQ schemes in a frame. As an example, in our SNVQ[36,41] scheme, where 60% of frames invoked the 36-bit PVQ while 40% used 41-bit SVQ scheme, the average number of bits per frame becomes $0.6 \cdot (36) + 0.4 \cdot (41) = 38$ bits. 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.

Table 3 details the characteristics of two high-quality quantization schemes. The first configuration utilised a

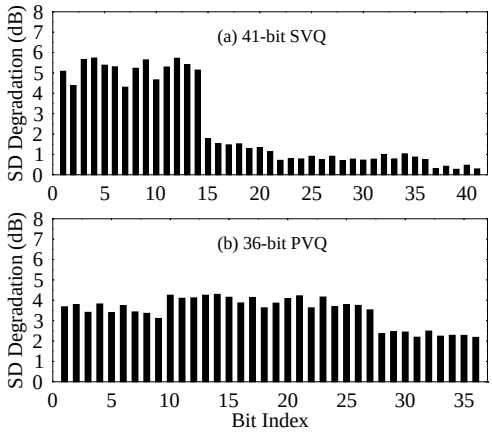


Figure 15: The SD degradations due to 100% Bit Error Rate in the (a) 41-bit SVQ and (b) 36-bit PVQ schemes.

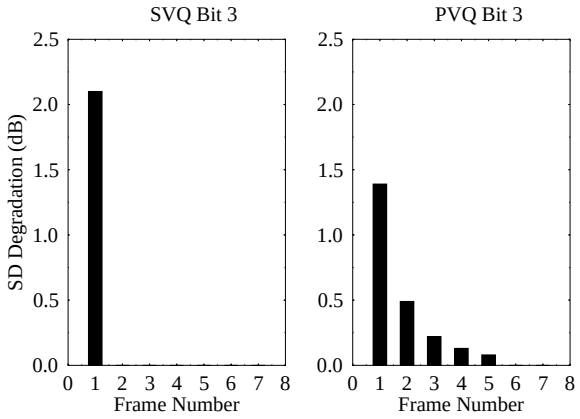


Figure 16: The SD degradation propagation effects for Bit 3 of 41-bit SVQ and 36-bit PVQ schemes, respectively.

$(4, 4, 4, 4)_{10,10,10,10}$ PVQ scheme employing 4·10= 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. Let us now consider the effects of transmission errors in the next section.

4 EFFECT OF CHANNEL ERRORS

In this section, we study the performance of the various LSF vector quantizers in the presence of channel errors. We first measured the bit sensitivity of the 41-bit SVQ and 36-bit PVQ schemes, which were employed in the SNVQ[36,41] arrangement. An often used approach to evaluating the bit sensitivity is to invert a given bit in every speech frame and measure the associated SD degradation inflicted. The error sensitivity of various bits for the 41-bit and 36-bit PVQ schemes are shown in Figure 15. Specifically, in the 41-bit SVQ scheme, 7 bits were used to quantize the 6-, 7- and 3-component subvectors, and their respective bit sensitivity is shown at bit indeces 1-21 in Figure 15(a). Additionally, a (4,4,4,4)-split second stage VQ was applied for further quantizing the quantization error due to the first-stage, using five bits for each sub-vector. This was plotted at bit indeces 21-42 in Figure 15(a), which exhibit a lower SD degradation compared to bit indeces 1-20. This is due to the higher bit sensitivity of those bits at the first stage, where the LSF subvectors were quantized, while at the second stage, the quantization error from the first stage - which is more noise-like - will be less sensitive to channel errors. For the 40-bit PVQ shown in Figure 15(b), almost all the bits exhibit similar SD degradation.

BER (%)	38-bit SNVQ	36-bit PVQ	41-bit SVQ
0	1.09	1.45	1.46
0.1	1.34	1.68	1.56
0.5	1.89	2.45	1.92
1	2.56	3.25	2.41
2	3.56	4.21	3.17
5	5.38	5.81	4.91

Table 4: SD comparisons of the 38-bit SNVQ, 36-bit PVQ and 41-bit SVQ for different Bit Error Rates (BER).

Using the method described above for quantifying the error sensitivity of various bits does not illustrate clearly the error propagation properties of different bits. In order to obtain a better picture of the error propagation effects, we employed another error sensitivity measure, where for each bit we find the average SD degradation due to a single bit error both in the frame in which the error occurs and in consecutive frames. These effects are shown in Figure 16 for bit index 3 of the 41-bit SVQ and 36-bit PVQ schemes, respectively. It can be observed that although the SVQ scheme exhibits a higher SD in Frame 1, the bit errors do not propagate over consecutive frames. By contrast, Bit 3 of the PVQ scheme in Figure 16 shows some error propagation effects due to the employment of the AR predictor, which uses recursive reconstruction of the LSFs.

In Table 4, we compare a 38-bit SNVQ with a 36-bit PVQ and 41-bit SVQ, using the SD measure for various bit error rates (BER). It can be observed that the Memoryless SVQ, which showed the best SD performance with increasing BERs. Note that the performance degrades further for the PVQ scheme, when the BER is increased, compared to the SNVQ and SVQ schemes. In the forthcoming section we provide a brief subjective evaluation of our previous schemes.

Speech Material A	Speech Material B	Preference		
		A (%)	B (%)	Neither (%)
SNVQ[36,41] BER 0% - UQR	41-bit SVQ BER 0% - UQR	75.0	0.0	25.0
SNVQ[36,41] BER 1% - UQR	41-bit SVQ BER 1% - UQR	55.0	10.0	35.0
SNVQ[36,41] BER 0% - QR	41-bit SVQ BER 0% - QR	50.0	30.0	20.0
SNVQ[36,41] BER 1% - QR	41-bit SVQ BER 1% - QR	25.0	15.0	60.0

Table 5: Details of the listening tests conducted using the pairwise comparison method, where the listeners were given a choice of preference between two speech files coded using SNVQ[36,41] or 41-bit SVQ schemes. UQR denotes using unquantized residual while QR means using the quantized residual by ACELP.

5 SUBJECTIVE EVALUATION

Informal listening tests were conducted, in order to verify the objective results obtained in the previous section. In the test, the 38-bit SNVQ[36,41] scheme was compared to the 41-bit Memoryless SVQ benchmarker. The coders were compared both for noiseless conditions and for a BER of 1%. We used the model shown in Figure 3 to evaluate the perceptual speech quality after applying LSF vector quantization. The prediction residual was formed by filtering the speech signal using the unquantized prediction filter coefficients, and the reconstructed speech was generated by exciting a quantized synthesis filter with the undistorted residual. In this way, the effects of LSF quantization can be studied separately from the encoding of the residual.

For the sake of completeness, we also evaluated the subjective performance of the vector quantizers in conjunction with a wideband ACELP codec. In the ACELP codec, we performed the LPC analysis every 10 ms and the pitch analysis for every subframe of 5 ms, in the form of an adaptive codebook. The adaptive codebook index and gain were quantized using 8 bits and 5 bits, respectively. In every 5ms subframe, the process of fixed codebook search procedure was applied twice on the basis of two 2.5ms segments, which was found to reduce the complexity significantly without degrading the performance of the codec [21]. A 20-bit algebraic codebook was employed, which consists of 15 bits encoding the five excitation pulse positions and an additional 5 bits to encode the sign of each pulse. Thus the bit rate used for the quantization of LSFs employing the SNVQ[36,41] scheme was 3.8 kbit/s, and that used for the quantization of the adaptive- and fixed-codebook parameters was 12.2 kbit/s. This gave a total bit rate of 16 kbit/s.

The subjective evaluation was carried out here through pairwise comparison tests using 10 listeners. Eight sentences spoken by four males and four females were used for the evaluation. The sentences were pairwise compared, including some comparisons with the uncoded original sentences. All possible A-B pairs were generated and presented in a randomized order. The listeners' task was to indicate their preference concerning either one or the other of the coded versions, or to indicate no preference. Our results accruing from these informal tests were shown in Table 5. The listening tests revealed that under perfect channel conditions, using the unquantized residual (UQR), the SNVQ was preferred to the 41-bit SVQ scheme in 75% of the comparisons. For a channel exhibiting 1% BER, 55% of the listeners favoured the SNVQ[36,41] scheme as shown in Table 5. In the case of quantizing the residual (QR) using the above ACELP codec under perfect channel conditions, the SNVQ[36,41] scheme still outperformed the 41-bit SVQ benchmarker, although at a BER of 1% its subjective superiority eroded.

6 CONCLUSIONS

In this contribution, we have comparatively studied various predictive and memoryless vector quantizers. In the context of memoryless vector quantization, 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 order to improve the performance of this initial memoryless scheme, we introduced V/UV classification. This approach gave about 0.2dB SD improvement, but increased the complexity. Nonetheless, both of these sub-optimum approaches maintained a low computational complexity, as well as a high error resilience.

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 5, 10 and 13, 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.

7 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.

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 J. 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] NTIS Federal Computer Products Center, "DARPA TIMIT Acoustic-phonetic continuous speech corpus", US Department of Commerce.

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

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

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

[19] T. Eriksson, J. Linden and J. Skoglund, "Interframe LSF Quantization for Noisy Channels," *IEEE Transactions on Speech and Audio Processing*, vol. 7, pp. 495–509, Sept 1999.

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

[21] J.W. Paulus and J. Schnitzler, "16kbit/s Wideband Speech Coding Based On Unequal Subbands," in *Proceedings of ICASSP*, pp. 255–258, 1996.