Switched Split Vector Quantisation of Line Spectral Frequencies for Wideband Speech Coding

Stephen So and Kuldip K. Paliwal

School of Microelectronic Engineering,
Griffith University, Brisbane, Australia, 4111.
s.so@griffith.edu.au, k.paliwal@griffith.edu.au

Abstract

In this paper, we investigate the use of the switched split vector quantiser (SSVQ) for coding short-term spectral envelope information for wideband speech coding. The SSVQ is the hybrid of a switch vector quantiser and split vector quantiser, which has been shown in previous studies to be more efficient, in terms of rate-distortion, as well as possessing low computational complexity, than the split vector quantiser. In our experiments, the SSVQ is used to quantise line spectral frequencies from the TIMIT database and its spectral distortion performance is compared with the split vector quantiser, the split-multistage vector quantiser (S-MSVQ) with MA predictor from the AMR-WB speech coder (ITU-T G.722.2), and PDF-optimised scalar quantisers. We show the SSVQ, which is a memoryless scheme, to achieve comparable spectral distortion to the S-MSVQ with MA predictor at 46 bits/frame. The five-part SSVQ requires 42 bits/frame and 17.7 kflops/frame for transparent coding, compared with 46 bits/frame and 40.96 kflops/frame for the five-part split vector quantiser.

1. Introduction

The quantisation of linear predictive coding (LPC) parameters in CELP coders for narrowband speech (300–3400 Hz) has been thoroughly investigated in the literature, where product code vector quantisers operating on vectors of 10 line spectral frequency (LSF) parameters [10], generally require 24 bits/frame for transparent quality [15, 11]. With the introduction of high-speed data services in wireless communication systems, wideband speech (50–7000 Hz) can now be accommodated [2]. Wideband speech has improved naturalness and intelligibility due to the added bandwidth. However, wideband CELP coders typically require 16 LPC parameters for representing the speech spectral envelope, hence vector quantisers need to operate at higher bitrates and on vectors of larger dimension.

Harborg *et al.* [9] quantised 16 to 18 log-area-ratio coefficients at 60 to 80 bits/frame using non-uniform scalar quantisers. Lefebvre *et al.* [12] and Chen *et al.* [5] used a seven-part split vector quantiser operating at 49 bits/frame to quantise 16 LSF parameters. Transparent results were reported by Biundo *et al.* [4] for a four and five part split vector quantiser at 45 bits/frame. Because successive LSF frames are highly correlated [8], better quantisation can be achieved by exploiting the interframe correlation. Ubale and Gersho [21] used a seven-stage tree-searched multistage vector quantiser [11] with a moving average (MA) predictor at 28 bits/frame, while Biundo *et al.* [4] reported transparent results using an MA predictive split-multistage vector quantiser (S-MSVQ) at 42 bits/frame. Guibé *et al.* [8] achieved transparent coding using a safety-net vector quantiser at 38 bits/frame, while the Adaptive Multi-Rate wideband (AMR-WB) speech codec [2, 1] uses an S-MSVQ with MA predictor at 46 bits/frame. Other quantisation schemes recently reported include the predictive Trellis-coded quantiser [16], the HMM-based recursive quantiser [6], and the multi-frame GMM-based block quantiser [19], which achieve a spectral distortion of 1 dB at 34, 40, and 37 bits/frame, respectively.

In our previous study [18], we showed how the losses in the shape and memory advantages [14] incurred by the split vector quantiser, are compensated by the switched split vector quantiser (SSVQ), which result in better rate-distortion performance for narrowband LSF quantisation [17]. Another characteristic of SSVQ is the low computational complexity, which comes at the expense of an increase in memory requirements. In this paper, we examine the performance of the SSVQ on LSFs from wideband speech and compare its rate-distortion performance with the split vector quantiser (SVQ), PDF-optimised scalar quantisers, and the split-multistage vector quantiser (S-MSVQ) with moving average (MA) predictor from the AMR-WB speech codec (ITU-T G.722.2). We show that the SSVQ, which is a memoryless quantisation scheme, outperforms SVQ and achieves comparable spectral distortion with the S-MSVQ with MA predictor at 46 bits/frame.

2. Switched split vector quantisation

The basic idea of SSVQ is to populate the vector space with many local split vector quantisers, while switching to one of them based on a nearest-neighbour criterion and quantising the vector using the respective codebook. Correlation that exists across all dimensions of the vector space can be exploited as these local SVQs are positioned via an optimal vector quantiser, which we refer to as the *switch vector quantiser*, that is designed using the Linde-Buzo-Gray (LBG) algorithm [13] on all the vectors. Furthermore, this positioning of local SVQs via the LBG algorithm allows for a better matching of the source probability density function (PDF) shape [18]. For each local SVQ, the 16-dimensional LSF vector is split into five parts with (3, 3, 3, 3, 4) division, as is done in [4]. Bits are uniformly distributed to each part where-ever possible, with preference given to higher frequency LSFs, when the number of bits is not divisible by five.

2.1. LSF representation and weighted distance measure

It is common practice in speech coders to quantise the line spectral frequency (LSF) representation [10] of the linear prediction (LP) coefficients, as they possess desirable qualities such as localisation in frequency of quantisation errors and simple verification of synthesis filter stability [15, 20]. SVQs are more suit-

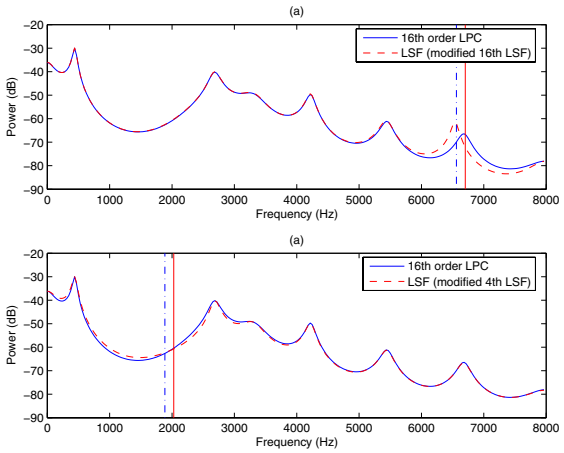

Figure 1: Original and reconstructed spectral envelope for a 16th order LPC analysis: (a) shifting the 15th LSF (SD=1.8515 dB); (b) shifting the 4th LSF (SD=0.7548 dB). The solid and dashed vertical lines show the original and shifted LSFs, respectively.

able than multistage vector quantisers (MSVQ) for quantising the LSF vectors as they can easily ensure correct LSF ordering for the filter stability.

We also introduce a modified form of the weighted distance measure from [15], which places more emphasis on the LSFs representing the formant regions. This is because, as shown in Fig. 1, deviations in the LSFs situated in higher power regions affect the reconstructed power spectrum more than those in the spectral valleys. The weighted distance measure between the original vector, $\boldsymbol{f}$, and the approximated vector, $\hat{\boldsymbol{f}}$, is defined as:

$$d_w(\boldsymbol{f}, \hat{\boldsymbol{f}}) = \sum_{i=1}^{16} \left[w_i (f_i - \hat{f}_i) \right]^2 \quad (1)$$

where f_i and $\hat{f}_i$ are the ith LSF in the original and approximated vector respectively. The dynamic weights, $\{w_i\}_{i=1}^{16}$, are given by:

$$w_i = [P(f_i)]^r \quad (2)$$

where $P(f)$ is the LPC power spectrum and r is a constant (typical value used is 0.15) [15].

2.2. SSVQ codebook training

Fig. 2 shows a block diagram of the SSVQ codebook training. The LBG algorithm [7] is first applied on all vectors to produce m centroids (or means) $\{\boldsymbol{\mu}_i\}_{i=1}^{m}$. In the Euclidean distortion sense, these centroids are the 'best' representation of all the vectors in that Voronoi region. Hence, we can use them to form the *switch VQ codebook* which will be used for switch-direction selection. All the training vectors are classified based on the nearest-neighbour criterion:

$$j = \underset{i}{\operatorname{argmin}}\, d_w(\boldsymbol{x}, \boldsymbol{\mu}_i) \quad (3)$$

where $\boldsymbol{x}$ is the vector under consideration and j is the cluster (or, switching direction) to which the vector is classified. With

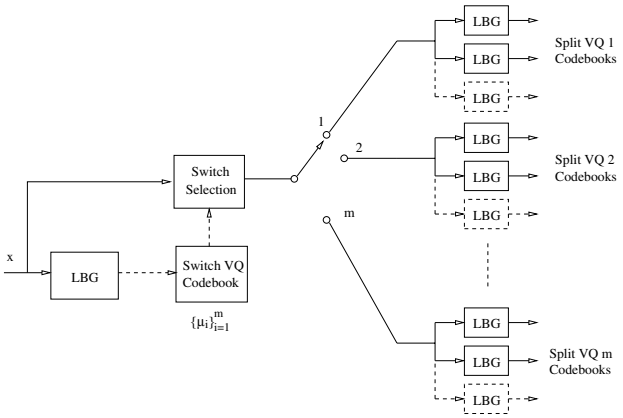

Figure 2: SSVQ Codebook Training

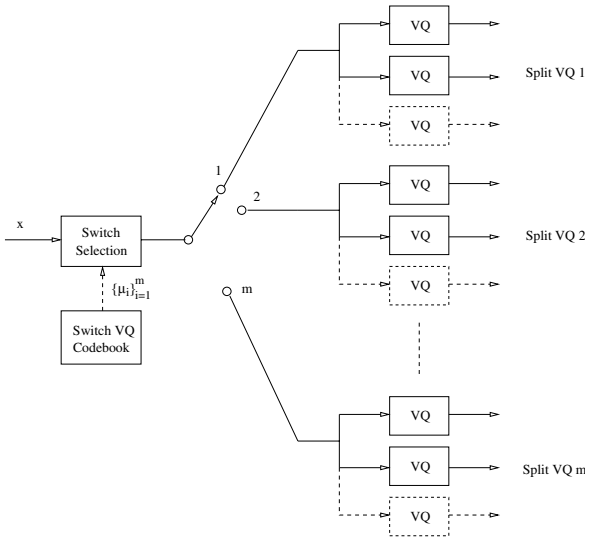

Figure 3: SSVQ Coding

the training vectors classified to the m clusters, local SVQ codebooks are designed for each cluster (or, switching direction) using the corresponding training vectors.

2.3. SSVQ coding

Fig. 3 shows a block diagram of SSVQ coding. Each vector to be quantised is first switched to one of the m possible directions based on the nearest-neighbour criterion defined by (3), using the switch VQ codebook, $\{\boldsymbol{\mu}_i\}_{i=1}^{m}$, and then quantised using the corresponding SVQ.

3. Distortion measures for LPC parameters

In order to objectively measure the distortion between a coded and uncoded LPC parameter vector, the spectral distortion is often used in narrowband speech coding [15]. For the ith frame,

Table 1: Average spectral distortion (SD), computational complexity, and memory requirements (ROM) of the five-part switched split vector quantiser as a function of bitrate and number of switch directions on wideband LSF vectors from the TIMIT database

m	Bits/ frame	Avg. SD (in dB)	Outliers (in %) 2–4 dB	Outliers (in %) >4 dB	kflops/ frame	ROM (floats)
8	46	0.889	0.33	0.00	27.1	53376
	45	0.922	0.43	0.00	24.1	47232
	44	0.953	0.57	0.00	21.0	41088
	43	0.986	0.66	0.00	19.5	38016
	42	1.037	1.05	0.00	15.4	34944
16	46	0.878	0.34	0.00	24.6	94464
	45	0.906	0.44	0.00	21.5	82176
	44	0.936	0.50	0.00	20.0	76032
	43	0.975	0.64	0.00	18.4	69888
	42	1.018	0.83	0.00	17.7	66816

the spectral distortion (in dB), D_i, is defined as:

$$D_i = \sqrt{\frac{1}{F_s}\int_0^{F_s} \left[10\log_{10} P_i(f) - 10\log_{10} \hat{P}_i(f)\right]^2 df} \quad (4)$$

where F_s is the sampling frequency and $P_i(f)$ and $\hat{P}_i(f)$ are the LPC power spectra of the coded and uncoded ith frame, respectively. The conditions for transparent speech from narrowband LPC parameter quantisation are [15]:

1. The average spectral distortion (SD) is approximately 1 dB,
2. there is no outlier frame having more than 4 dB of spectral distortion, and
3. less than 2% of outlier frames are within the range of 2–4 dB.

According to Guibé *et al.* [8], listening tests have shown that these conditions for transparency, which are often quoted in the narrowband speech coding literature, also apply to the wideband case.

4. Experimental setup

The TIMIT database was used in the training and testing of the SSVQ, where speech is sampled at 16 kHz. We have used the preprocessing and LPC analysis of the AMR-WB speech codec (floating point version) [1] to produce linear prediction coefficients which are then converted to line spectral frequency (LSF) representation [10]. The training set consists of 333789 vectors while the evaluation set, consisting of speech not contained in the training, has 85353 vectors.

We have also tested the split-multistage vector quantiser (S-MSVQ) from the AMR-WB speech codec on the database, so that it can be used for comparison. Immittance spectral pairs (ISP) [3] are used in the AMR-WB codec while the quantisation scheme considered in this paper quantises line spectral frequencies (LSF). This presents no problem in our spectral distortion comparisons as (4) requires linear predictive coefficients, which can be obtained from ISPs and LSFs.

As a further basis for comparison, we have applied scalar quantisers with non-uniform bit allocation to code wideband LSF vectors. Each scalar quantiser is designed using the generalised Lloyd algorithm [7] and bit allocation is performed in a way that is similar to the one presented in [20].

Table 2: Average spectral distortion (SD), computational complexity, and memory requirements (ROM) of the five-part split vector quantiser as a function of bitrate on wideband LSF vectors from the TIMIT database

Bits/frame	Avg. SD (in dB)	Outliers (in %) 2–4 dB	Outliers (in %) >4 dB	kflops/ frame	ROM (floats)
46	1.012	0.68	0.00	40.96	10240
45	1.061	0.99	0.00	32.76	8192
44	1.092	1.10	0.00	29.69	7424
43	1.151	1.70	0.00	26.62	6656
42	1.200	2.31	0.00	23.55	5888

Table 3: Average spectral distortion as a function of bitrate of the split-multistage vector quantiser with MA predictor in AMR-WB speech codec on wideband LSF vectors from the TIMIT database

Bits/frame	Avg. SD (dB)	Outliers (in %) 2–4 dB	Outliers (in %) >4 dB
46	0.894	0.76	0.01
36	1.304	5.94	0.03

5. Results and discussion

Table 1 shows the average spectral distortion, computational complexity, and memory requirements of the five-part SSVQ at varying bitrates and number of switch directions. We can see that by increasing the number of switch directions from 8 to 16, lower spectral distortion is achieved at all bitrates. This is attributed to the better exploitation of global vector correlation and matching of the PDF shape by the switch vector quantiser [18]. Transparent coding has been achieved at 42 bits/frame.

Table 2 shows the average spectral distortion, computational complexity, and memory requirements of a five-part split vector quantiser, which uses unweighted mean squared error. It uses the same partition sizes as the five-part SSVQ. We can see that the five-part SVQ requires 46 bits/frame to achieve transparent coding. Comparing these results with Table 1, we observe a saving of up to 4 bits/frame for transparent coding with the SSVQ. Also, the computational complexity of the transparent SSVQ is less than 40% of the complexity of the transparent SVQ. This confirms the better rate-distortion and computational efficiency of the SSVQ over the SVQ, due to the former's compensation of the memory and shape advantage losses of the latter [18].

Table 3 shows the average spectral distortion of the S-MSVQ with MA predictor. Comparing this Table with Table 1, we can see that the SSVQ, which is a memoryless scheme, achieves comparable spectral distortion performance to the S-MSVQ with MA predictor scheme at 46 bits/frame. In addition, the SSVQ has produced only half the number of outlier frames than the S-MSVQ, which is to be expected, since the latter has a predictive component [8].

Table 4 shows the average spectral distortion performance of the PDF-optimised scalar quantisers. To achieve a spectral

Table 4: Average spectral distortion of the PDF-optimised scalar quantisers as a function of bitrate on wideband LSF vectors from the TIMIT database

Bits/frame	Avg. SD (dB)	Outliers (in %)	
		2–4 dB	> 4 dB
60	0.970	0.95	0.00
59	1.011	1.18	0.01
58	1.080	1.64	0.01
57	1.120	1.88	0.01

distortion of approximately 1 dB, we need about 59 bits/frame. Comparing this with Table 1, we can see that the SSVQ requires 17 bits/frame less than PDF-optimised scalar quantisers to achieve 1 dB spectral distortion.

6. Conclusion

In this paper, we have investigated the use of the switched split vector quantiser and applied it to the problem of coding line spectral frequencies for wideband speech. In our LSF quantisation experiments on the TIMIT database, we have shown the SSVQ to provide a better trade-off between bitrate and distortion performance than the traditional SVQ. That is, for a given bitrate, the SSVQ not only gives less spectral distortion than the SVQ, but it also reduces the computational (search) complexity. These advantages come at the cost of increased memory requirements. We have also evaluated the S-MSVQ with MA predictor from the AMR-WB speech coder and shown that the SSVQ, which by comparison is a memoryless quantisation scheme, achieves comparable spectral distortion at 46 bits/frame. This suggests that further gains may be achieved by adding a memory component to the SSVQ.

7. References

[1] "3rd generation partnership project; Technical specification group services and system aspects; Speech codec speech processing functions; AMR wideband speech codec; Transcoding functions", 3GPP TS 26.190.

[2] B. Bessette, R. Salami, R. Lefebvre, M. Jelínek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Järvinen, "The adaptive multirate wideband speech codec (AMR-WB)", *IEEE Trans. Speech Audio Processing*, vol. 10, no. 8, pp. 620–636, Nov. 2002.

[3] Y. Bistritz and S. Pellerm, "Immittance spectral pairs (ISP) for speech encoding", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1993, pp. II-9–II-12.

[4] G. Biundo, S. Grassi, M. Ansorge, F. Pellandini and P.A. Farine, "Design techniques for spectral quantization in wideband speech coding", in *Proc. of 3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, Budapest, Oct. 2002, pp. 114-119.

[5] J.H. Chen and D. Wang, "Transform predictive coding of wideband speech signals", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1996, pp. 275–278.

[6] E.R. Duni, A.D. Subramaniam, and B.D. Rao, "Improved quantization structures using generalised HMM modelling with application to wideband speech coding", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 2004, pp. 161-164.

[7] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Boston: Kluwer, 1991.

[8] G. Guibé, H.T. How and L. Hanzo, "Speech spectral quantizers for wideband speech coding", *European Transactions on Telecommunications*, 12(6), pp. 535–545, 2001.

[9] E. Harborg, J.E. Knudsen, A. Fuldseth and F.T. Johansen, "A real-time wideband CELP coder for a videophone application", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1994, pp. 121–124.

[10] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals", *J. Acoust. Soc. Amer.*, vol. 57, p. S35, Apr. 1975.

[11] W.P. LeBlanc, B. Bhattarchaya, S.A. Mahmoud and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding", *IEEE Trans. Speech Audio Processing*, Vol. 1, pp. 373–385, Oct. 1993.

[12] R. Lefebvre, R. Salami, C. Laflamme, J.P. Adoul, "High quality coding of wideband audio signals using transform coded excitation (TCX)", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1994, pp. 193–196.

[13] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design", *IEEE Trans. Commun.*, vol. COM-28, no. 1, pp. 84–95, Jan. 1980.

[14] T.D. Lookabaugh and R.M. Gray, "High-resolution quantization theory and the vector quantizer advantage", *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1020–1033, Sept 1989.

[15] K.K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", *IEEE Trans. Speech Audio Processing*, Vol. 1, No. 1, pp. 3–14, Jan. 1993.

[16] Y. Shin, S. Kang, T.R. Fischer, C. Son, and Y. Lee, "Low-complexity predictive trellis coded quantization of wideband speech LSF parameters", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2004, pp. 145-148.

[17] S. So and K.K. Paliwal, "Efficient vector quantisation of line spectral frequencies using the switched split vector quantiser", in *Proc. Int. Conf. Spoken Language Processing*, Jeju, Korea, Oct. 2004.

[18] S. So and K.K. Paliwal, "Efficient product code vector quantisation using the switched split vector quantiser", submitted to *Digital Signal Processing*, 2004.

[19] S. So and K.K. Paliwal, "Multi-frame GMM-based block quantisation of line spectral frequencies for wideband speech coding", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Philadelphia, USA, 2005.

[20] F.K. Soong and B.H. Juang, "Line spectrum pair (LSP) and speech data compression", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Diego, California, Mar 1984, pp. 37–40.

[21] A. Ubale and A. Gersho, "A multi-band CELP wideband speech coder", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1994, pp. 1367-1370.