A COMPARISON OF LSF AND ISP REPRESENTATIONS FOR WIDEBAND LPC PARAMETER CODING USING THE SWITCHED SPLIT VECTOR QUANTISER

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 apply the switched split vector quantiser (SSVQ) for coding short-term spectral envelope information in wideband speech coding to compare and contrast two LPC parameter representations: line spectral frequencies (LSFs) and immittance spectral pairs (ISPs). 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 lower computational complexity, than the split vector quantiser (SVQ). On the TIMIT database, the five-part SSVQ requires 43 and 44 bits/frame to transparently code LSFs and ISPs, respectively. This 1 bit/frame difference between LSFs and ISPs is also observed in five-part SVQ. The split-multistage vector quantiser (S-MSVQ) with MA predictor from the AMR-WB speech coder (ITU-T G.722.2) is also used as a basis of comparison where we find the SSVQ, which is a memoryless quantisation scheme, to be competitive at 46 bits/frame.

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 [9], generally require 24 bits/frame for transparent quality [14, 10]. 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.* [8] quantised 16 to 18 log-area-ratio coefficients at 60 to 80 bits/frame using non-uniform scalar quantisers. Lefebvre *et al.* [11] 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 [7], better quantisation can be achieved by exploiting the interframe correlation. Ubale and Gersho [20] used a seven-stage tree-searched multistage vector quantiser [10] 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.* [7] achieved transparent coding us-

ing 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 [15], the HMM-based recursive quantiser [6], and the multi-frame GMM-based block quantiser [18], which achieve a spectral distortion of 1 dB at 34, 40, and 37 bits/frame, respectively.

So and Paliwal [17] showed how the losses in the shape and memory advantages [13] 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 [16]. Another characteristic of SSVQ is the low computational complexity, which comes at the expense of an increase in memory requirements. In this paper, we evaluate the performance of the SSVQ on the two popular LPC parameter representations used in wideband speech coders: line spectral frequencies (LSFs) [9] and immittance spectral pairs[1] (ISPs) [3]. We also compare the SSVQ with the split vector quantiser (SVQ) and split-multistage vector quantiser (S-MSVQ) with moving average (MA) predictor from the AMR-WB speech coder.

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 [12] 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 [17]. For each local SVQ, the 16-dimensional LPC parameter 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 or ISPs, when the number of bits is not divisible by five.

2.1. The Line Spectral Frequency and Immittance Spectral Pair Representation

Most narrowband speech coders, such as the FS-1016 4.8 kbps CELP coder and AMR speech coder, use the *line spectral fre-*

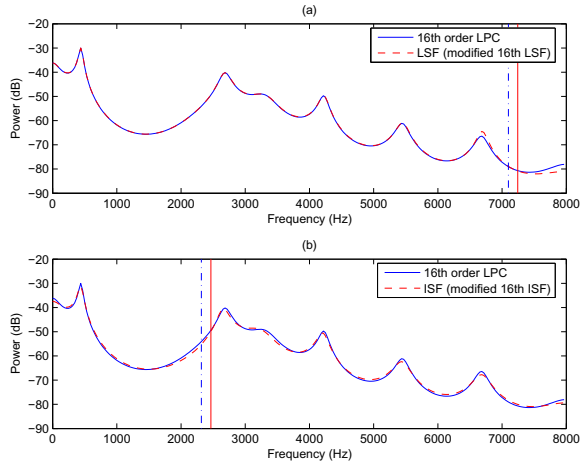[1]ISPs are used in the AMR-WB (ITU-T G.722.2) speech coder.

Figure 1: Original and reconstructed power spectral envelope estimates for 16th order LPC analysis: (a) Shifting the 16th LSF by 142 Hz (SD=0.583 dB); (b) Shifting the 16th ISF by 142 Hz (SD=0.684 dB). The solid and dashed vertical lines show the original and shifted parameters (LSF and ISF), respectively.

quency (LSF) representation [9] for representing the short-term spectral envelope. LSFs are quantised instead 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 [14, 19].

The line spectral frequencies are defined as the roots of the following polynomials:

$$P(z) = A(z) + z^{-(p+1)}A(z^{-1}) \qquad (1)$$

and

$$Q(z) = A(z) - z^{-(p+1)}A(z^{-1}) \qquad (2)$$

where p is the order of the LPC analysis and $A(z)$ is the LPC synthesis filter. These two polynomials, $P(z)$ and $Q(z)$, are parametric models of the acoustic tube in two extremal states, where the $(p+1)$th stage (representing the glottis) is either completely closed or completely opened, respectively [19]. Consequently, LSFs have the following properties [19]:

1. All zeros of $P(z)$ and $Q(z)$ lie on the unit circle;
2. zeros of $P(z)$ and $Q(z)$ are interlaced with each other; and
3. the minimum phase property of $A(z)$ is easily preserved after quantisation of the LSFs if the first two properties are satisfied.

Therefore, p LPC coefficients, $[a_1, a_2, \ldots, a_p]$, can be converted to p line spectral frequencies, $[\omega_1, \omega_2, \ldots, \omega_p]$.

The *immittance spectral pairs* (ISP) representation was introduced by Bistritz and Peller [3]. It consists of the poles and zeros of the following immittance function at the glottis [3]:

$$\mathcal{I}_p(z) = \frac{A(z) - z^{-p}A(z^{-1})}{A(z) + z^{-p}A(z^{-1})} \qquad (3)$$

as well as the pth reflection coefficient, k_p. The poles and zeros of the immittance function possess the same properties as the LSFs.
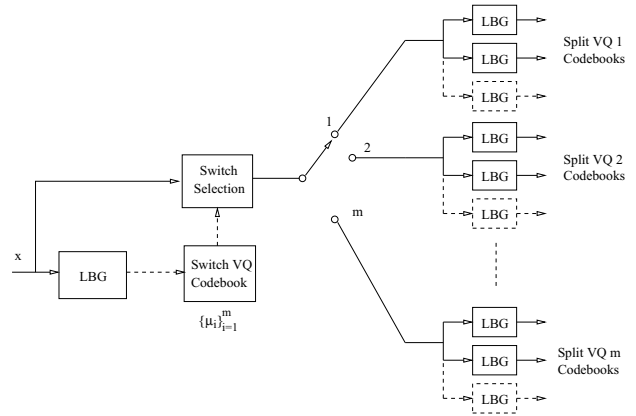

Figure 2: SSVQ Codebook Training

Therefore, p LPC coefficients, $[a_1, a_2, \ldots, a_p]$, can be converted to $p - 1$ immittance spectral pairs and a reflection coefficient, $[\cos \omega_1, \cos \omega_2, \ldots, \cos \omega_{p-1}, k_p]$ [3]. Because the reflection coefficient is a different variable from the first $p - 1$ 'frequencies' (since they lie on the unit circle), it possesses different quantisation sensitivities, as shown in Fig. 1. The AMR-WB speech coder quantises the frequency form of ISPs, which are also known as the *immittance spectral frequencies* (ISFs). ISFs are expressed as $[\omega_1, \omega_2, \ldots, \omega_{p-1}, \frac{1}{2}\cos^{-1} k_p]$ [1]. The arc-cosine tends to flatten the sensitivity curve of the pth reflection coefficient.

2.2. SSVQ codebook training

Fig. 2 shows a block diagram of the SSVQ codebook training. The LBG algorithm [12] 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:

$$j = \operatorname*{argmin}_i d(\boldsymbol{x}, \boldsymbol{\mu}_i) \qquad (4)$$

where $\boldsymbol{x}$ is the vector under consideration, j is the cluster (or, switching direction) to which the vector is classified, and $d(\boldsymbol{x}, \hat{\boldsymbol{x}})$ is the mean squared error between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$. With 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 (4), 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
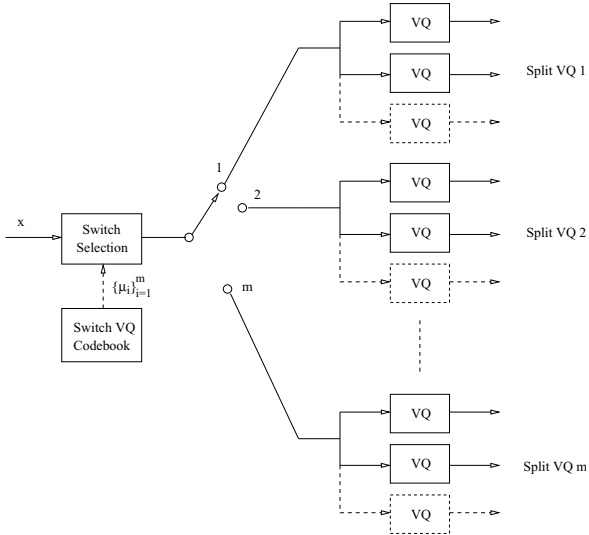
Figure 3: SSVQ Coding

in narrowband speech coding [14]. For the ith frame, the spectral distortion (in dB), D_i, is defined as:

$$D_i = \sqrt{\frac{1}{F_s} \int_0^{F_s} \left[10 \log_{10} P_i(f) - 10 \log_{10} \hat{P}_i(f) \right]^2 df} \quad (5)$$

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 [14]:

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.* [7], 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 LSFs and ISFs. The training set consists of 333789 vectors while the evaluation set, which consists 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.

5. RESULTS AND DISCUSSION

Table 1 shows the average spectral distortion, computational complexity, and memory requirements of the five-part SSVQ at vary-

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 (top half) and ISF (bottom half) vectors from the TIMIT database

m	Bits/ frame	Avg. SD (in dB)	Outliers (in %)		kflops/ frame	ROM (floats)
			2–4 dB	> 4 dB		
8	46	0.919	0.54	0.00	27.1	53376
	45	0.953	0.64	0.00	24.1	47232
	44	0.984	0.79	0.00	21.0	41088
	43	1.018	0.90	0.00	19.5	38016
	42	1.066	1.37	0.00	15.4	34944
16	46	0.903	0.48	0.00	24.6	94464
	45	0.932	0.60	0.00	21.5	82176
	44	0.964	0.73	0.00	20.0	76032
	43	1.007	0.97	0.00	18.4	69888
	42	1.050	1.19	0.01	17.7	66816
8	46	0.931	0.53	0.00	27.1	53376
	45	0.968	0.86	0.00	24.1	47232
	44	0.999	1.07	0.00	21.0	41088
	43	1.037	1.21	0.00	19.5	38016
	42	1.080	1.68	0.00	15.4	34944
16	46	0.920	0.63	0.00	24.6	94464
	45	0.948	0.77	0.00	21.5	82176
	44	0.983	0.87	0.00	20.0	76032
	43	1.032	1.30	0.00	18.4	69888
	42	1.078	1.57	0.01	17.7	66816

ing bitrates and number of switch directions. The top half of the Table are for LSFs and the bottom half are for ISFs. We can see that by increasing the number of switch directions from 8 to 16, lower spectral distortion is achieved at all bitrates. Also, we note that the spectral distortion incurred when quantising LSFs is lower than that when quantising ISFs. The number of outlier frames is also higher with ISFs. Transparent coding has been achieved at 43 bits/frame for LSFs and 44 bits/frame for ISFs.

Table 2 shows the average spectral distortion, computational complexity, and memory requirements of a five-part split vector quantiser. The same vector partition sizes were used. We can see that the five-part SVQ requires 46 bits/frame to achieve transparent coding for LSFs and 47 bits/frame for ISFs. Again, we observe better performance when quantising LSFs, amounting to a 1 bit/frame difference. Comparing these results with Table 1, we observe a saving of up to 4 bits/frame for transparent coding with the SSVQ over the SVQ. Also, the computational complexity of the transparent SSVQ is less than 40% of the complexity of the transparent SVQ.

Table 3 shows the average spectral distortion of the S-MSVQ with MA predictor. Comparing this Table with the bottom half of Table 1, we can see that the SSVQ achieves a slightly higher spectral distortion than the S-MSVQ with MA predictor scheme at 46 bits/frame. This is to be expected as SSVQ does not exploit interframe correlation. However, the SSVQ has produced less outlier frames than the S-MSVQ with MA predictor, since it does not have a predictive component, which tends to produce more outliers [7].

6. CONCLUSION AND FUTURE WORK

In this paper, we have applied the switched split vector quantiser to compare line spectral frequencies and immittance spectral pairs

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 (top half) and ISF (bottom half) vectors from the TIMIT database

Bits/frame	Avg. SD (in dB)	Outliers (in %)		kflops/ frame	ROM (floats)
		2–4 dB	> 4 dB		
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
47	0.997	0.70	0.00	47.10	11776
46	1.030	0.88	0.00	40.96	10240
45	1.070	1.21	0.00	32.76	8192
44	1.106	1.32	0.00	29.69	7424
43	1.168	2.13	0.00	26.62	6656

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	> 4 dB
46	0.894	0.76	0.01
36	1.304	5.94	0.03

derived from wideband speech. Our results have indicated that vector quantising LSFs is superior to ISFs by about 1 bit/frame. This difference may be attributed to the reflection coefficient in the ISP representation, which has different quantisation properties, hence joint quantisation may not be optimal. An extended comparative study of LSFs and ISFs in joint block and vector quantisation will be presented in an upcoming paper. We have also shown that SSVQ is more efficient than SVQ in terms of rate-distortion and computational complexity, for wideband LPC parameter quantisation.

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] 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.

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.