BIMODAL QUANTIZATION OF WIDEBAND SPEECH SPECTRAL INFORMATION

Driss Guerchi
College of Information Technology, UAE University, 17555, Al Ain, U.A.E.

Keywords: Wideband CELP coding, Bimodal vector quantization, Low-rate LPC quantization.

Abstract: In this work we introduce an efficient method to reduce the coding rate of the spectral information in an algebraic code-excited linear prediction (ACELP) wideband codec. The Bimodal Vector Quantization (BMVQ) exploits the interframe correlation in spectral information to reduce the coding rate while maintaining high coded speech quality. In the BMVQ training phase, two codebooks are separately designed for voiced and unvoiced speech. For each speech frame, the optimal codebook for the search procedure is selected according to the interframe correlation of the spectral information.
The BMVQ was successfully implemented in an ACELP wideband coder. The objective and subjective performance were found to be comparable to that of the combination of the split vector quantization and multistage vector quantization at 2.3 kbit/s.

1 INTRODUCTION

Vector Quantization (VQ)is the most popular framework to quantize spectrum parameters in Code-Excited Linear Prediction (CELP) codecs. Several variants of the vector quantization technique have been developed to reduce further the coding rate of the Linear Predictive Coding (LPC) vectors while maintaining high coded-speech quality (Agiomyrgiannakis and Stylianou, 2007). For example, both the narrowband G.729 (Salami, et al., 1998) and wideband G.722.2 (ITU-T G.722.2, 2003) codec standards use a combination of Split Vector Quantization (SVQ) and Multistage Vector Quantization (MSVQ) to reduce the search algorithm complexity.

While the coding efficiency of vector quantization relies on its use of the intraframe correlation (that is the correlation between the parameters of the same LPC vector), split vector quantization sacrifices some of this correlation to decrease the computational complexity. In SVQ, an input vector is divided in two or more subvectors, which are coded individually using separate codebooks. This quantization method lowers the search complexity but at the expense of reduced speech quality. In order to reduce the CELP coding rate, VQ algorithms are often applied to the error between successive LPC vectors (after conversion to one of the LPC frequency representation, such as Line Spectral Frequency (LSF) in G.729, or Immitance Spectral Frequency (ISF) in G.722.2). Better LPC quantization is achieved since some of the interframe correlation of the spectral information is exploited. However, this approach is still not fully efficient since it neglects the interaction of the vocal tract shape with the vocal cords pitch in voiced speech. The suboptimality of the above methods is reflected in the disjoint operation of spectrum parameters quantization and pitch analysis.

The interfame error of the LPC vectors is not uniform. While its variance in voiced speech is too small, in unvoiced speech two consecutive LPC vectors show very week correlation.
In this paper, we introduce a new technique exploiting the interframe correlation of the spectral information to lower the spectrum coding rate. Our approach, which we name Bimodal Vector Quantization (BMVQ), codes separately the spectral information in voiced and unvoiced speech.

In the BMVQ training phase, two disjoint codebooks are individually populated from the spectral parameters of voiced and unvoiced speech, but only one codebook is used in the encoding of each frame spectrum. As a consequence, the LPC quantization process in the BMVQ technique is preceded by the selection of the appropriate codebook. This selection is based on the interframe correlation of the current

Guerchi D. (2008).
BIMODAL QUANTIZATION OF WIDEBAND SPEECH SPECTRAL INFORMATION.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 151-155
DOI: 10.5220/0001939201510155

and previous LPC vectors. This approach not only reduces the LPC coding rate but also produces high coded speech quality.

2 CLASSICAL LPC VECTOR QUANTIZATION

2.1 General-purpose Vector Quantization

Most of the recent speech coder standards use one of the various vector quantization algorithms to code the spectral information. In contrast to scalar quantization, VQ techniques reduce the coding rate in spite of an increase in search computation. The performance of a VQ method is function of the size of the codebook. A codebook with more entries certainly excels in modeling the spectral parameters. However, this improvement in speech quality is achieved to the detriment of an increase in coding rate and computational complexity.

For example, in the G.729 narrowband codec standard, a combination of Multistage VQ and Split VQ (MSVQ) is used to determine which 10-dimensional LSF vector (among all the LSF codebook entries) corresponds most closely to the current frame LSF vector. In the first stage of the search procedure, a 7-bit codebook is searched for the closest match to the difference between the input and predicted LSF vectors, while in the second stage two codebooks of 5 bits each are examined, for a total coding rate of 1.8 kbit/s. In the G.722.2 wideband coding standard (Bessete, et al., 2002), the same VQ technique, with slight modifications, is employed to code 16 ISF coefficients. A total of 46 bits per 20-ms frame is allocated to coding the input ISF vector for all the codec modes, except for the 6.60 kbit/s coder which searches the closed codewords among 832 entries (for a bit rate of 1.8 kbit/s). These bit allocations ignore the acoustical characteristics of voiced and unvoiced speech (Tamni, et al., 2005). The above vector quantization algorithms can be categorized as general-purpose quantization techniques since they are applied jointly to both voiced and unvoiced speech.

2.2 Shortcomings of Classical Vector Quantization Techniques

To code efficiently the LPC coefficients, one must employ separate codebooks for voiced and unvoiced frames. In both G.729 and G.722.2 standards, the error between the current and past frame LPC vectors is quantized using a combination of split and multistage vector quantization. While the amplitude of this error is too small in voiced speech, its magnitude and dynamic range for unvoiced speech are significantly higher. Figure 1 shows the squared error between consecutive ISF vectors in a wideband speech signal. Each of the G.729 and G.722.2 codecs quantize this error with the same quantizer for both voiced and unvoiced frames. This approach is certainly inefficient since it dos not exploit the high interframe correlation of the voiced speech spectrum. The variable-rate multimode wideband speech codec (Jelinek and Salami, 2007), which is based on a source-controlled coding paradigm, utilizes separate coding modes for different classes of speech. However, the spectral information is encoded using the same quantizer in all coding modes.

An obvious and better approach consists of quantizing the voiced and unvoiced spectrum information separately. The source-controlled quantization of the spectrum parameters will evidently provide higher coding performance. In (Guerchi, 2007) the interframe correlation of spectrum parameters is used to reduce the computational complexity by almost 30 % while maintaining the coding rate fixed. An alternative method of exploiting the high interframe correlation in voiced speech consists of using a smaller-size quantizer for this class of speech signal.

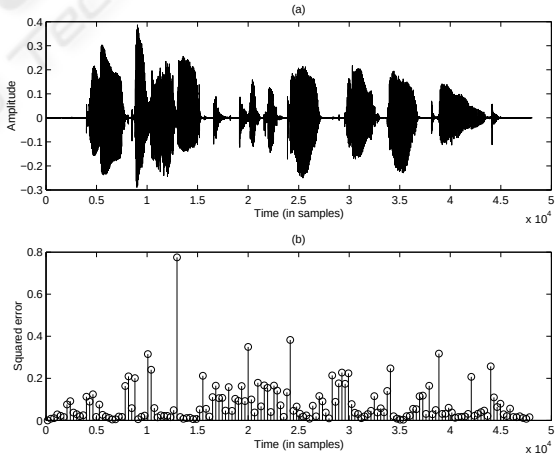

Figure 1: (a) Wideband speech signal; (b) ISF squared error between consecutive frames.

3 BIMODAL VECTOR QUANTIZATION

In this section, we introduce the bimodal vector quantization (BMVQ) technique. This technique, which consists of two disjoint ISF codebooks, reduces the

coding rate of the LPC coefficients while maintaining a toll speech quality. A voiced-speech ISF codebook (VCB) quantizes the spectral information in voiced speech. However, the BMVQ uses a separate ISF codebook (UCB) for unvoiced frames.
Each ISF input vector is quantized using exclusively either the VCB or UCB codebook. The two codebooks are trained individually from voiced and unvoiced speech segments.
The search of the closest match to an input LP filter vector is confined to only one codebook. The ISF vectors in steady-state voiced speech are highly correlated. The error between two consecutive ISF vectors is very small and its variance too. So the ISF error in voiced speech can be quantized with lesser amount of bits than in the traditional source-independent vector quantization techniques.
In the BMVQ approach, the LPC parameters are quantized differently according to the relative value of their interframe correlation. For each speech frame, linear prediction analysis is performed to extract 16 LPC coefficients. The estimated LPC vector is compared (after conversion to an equivalent ISF vector) to the quantized ISF vector of the past frame using a squared error distortion measure. The selection of the optimal codebook (VCB versus UCB) is based on the relative magnitude of this distortion. A small error distortion is a cue of quasi-stationary voiced speech.

We propose in this paper to exploit this interframe correlation to estimate the current ISF vector from the past frame ISF coefficients. The residual ISF vector r_n is quantized using either one of the BMVQ codebooks. The details of this algorithm are given in the next section. Figure 2 illustrates the procedure of the BMVQ algorithm.

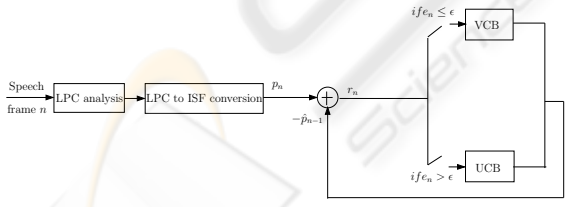

Figure 2: (a) Concept of the Bimodal Vector Quantization.

4 BIMODAL CODEBOOK DESIGN

In the BMVQ algorithm, two codebooks are trained from a large speech database. In the first phase of the training process, we manipulate the speech database to build two sets of speech segments. The first set contains voiced speech, while the second is populated from unvoiced speech.

Table 1: Bit Allocation of the 33-bit VCB.

Stage 1	stage 2
r_1 (6 bits)	$(r-\hat{r}_1)(0:2)$ (5 bits) $(r-\hat{r}_1)(3:5)$ (5 bits) $(r-\hat{r}_1)(6:8)$ (5 bits)
r_2 (6 bits)	$(r-\hat{r}_2)(0:2)$ (3 bits) $(r-\hat{r}_2)(3:6)$ (3 bits)

Table 2: Bit Allocation of the 33-bit UCB.

Stage 1	stage 2
r_1 (6 bits)	$(r-\hat{r}_1)(0:2)$ (5 bits) $(r-\hat{r}_1)(3:5)$ (4 bits) $(r-\hat{r}_1)(6:8)$ (4 bits)
r_2 (6 bits)	$(r-\hat{r}_2)(0:2)$ (4 bits) $(r-\hat{r}_2)(3:6)$ (4 bits)

In the second phase, linear predictive analysis of order 16 is performed on 20-ms speech frames from each set. The obtained LPC vectors from each speech class are first converted to ISF vectors and then used to design two ISF codebooks, VCB and UCB. We have adopted in the codebook design a combination of SVQ and MSVQ. The quantization procedure is similar to the one used in the G.722.2 standard but with lesser amount of bits.
It is worth noting that for the VCB codebook, a much smaller bit rate is sufficient to produce the same objective and subjective performance as that of the G.722.2 codec standard. The error between two consecutive LPC vectors is too small in voiced speech. Its variance allows significant bit-rate reduction.

At the end of the training process, each codebook will be characterized by one index l, $l = 0,1$. Tables 1 and 2 show the bit allocation of the VCB and UCB, respectively, where r_n is the ISF error vector between the current ISF vector, p_n, and the last frame quantized vector, $\hat{p}_{n-1}$.

In the first stage of the quantization process, the error vector r_n is split into two subvectors $r_{n,1}$ and $r_{n,2}$ of 9 and 7 coefficients, respectively. These two subvectors are quantized to $\hat{r}_{n,1}$ and $\hat{r}_{n,2}$. In the second stage, the resulting quantization errors $r_n - \hat{r}_{n,1}$ and $r_n - \hat{r}_{n,2}$ are split into three subvectors and two subvectors, respectively.

To achieve a speech quality comparable to that produced by the G.722.2 standard, our experiments confirmed that a total amount of 33 bits is required for voiced speech ISF codebook (ie, for VCB).

Even though, the energy of the error r_n is comparatively higher in unvoiced speech, the same amount of 33 bits (for the UCB) is sufficient to provide high

coded-speech quality since unvoiced speech is less sensitive to quantization errors.

One more bit is to be transmitted as an index of the selected codebook. The average coding rate for the spectral information is equal to 1.7 kbit/s (34 bits per 20-ms frame). This is a relative gain of 0.6 kbit/s compared to the SMVQ method in the G.722.2 standard.

5 SOURCE-CONTROLLED CODEBOOK

For every speech frame, the input ISF vector, p_n, is compared to the last frame quantized ISF vector $\hat{p}_{n-1}$. A comparator checks the error distortion, $r_n = p_n - \hat{p}_{n-1}$, between the two vectors. If the energy of r_n is smaller than a certain threshold ε, then the VCB will be used for the search of the closest codeword to the input vector p_n. Otherwise, the UCB will be selected as the optimal codebook for the ISF quantization. The selection of the best codebook (VCB versus UCB) is controlled by the type of the source signal (voiced versus unvoiced). The advantage of this source-controlled method is that for steady-state speech frames, the chance of hitting the optimum ISF vector in the VCB codebook is very high. Table 3 illustrates the algorithm for optimal codebook selection.

6 EVALUATION

We have conducted several simulations to compare the performance, in terms of objective and subjective measures, of the BMVQ technique to the G.722.2 SMVQ approach. The codebooks in both techniques have been trained using the same database. This is to avoid any effects of the selection of the database on the performance results. As an objective measure, we adopt the Segmental Signal-to-Noise Ratio (SegSNR) at the output of the decoder. The systems to be evaluated are two versions of the same wideband Algebraic CELP (ACELP) coder. The two coders are similar except in the ISF quantization, where in coder 1 we use the SMVQ approach. In the second ACELP coder, we implement the BMVQ algorithm. The database for the codebooks training consists of 150 minutes of English speech uttered by 8 speakers; four women and four men. Each speaker read the same short utterance 10 times. We used the squared error ISF distortion for training and testing. However, the weighted distortion measure of Paliwal and Atal (Paliwal and Atal, 1993) is used to evaluate the ISF quantization in both versions of the ACELP coder. The evaluation simulations have been conducted on six different input sentences uttered by other speakers. In Table 4, we present the SegSNR for both ACELP versions. Table 5 shows the average spectral distortion (Avg SD) between the input ISF vectors and their corresponding quantized versions. Informal listening tests have been performed as a subjective measures. In these comparative tests, for each speech signal, listeners have to listen to both signals produced by the BMVQ and SMVQ-based coders. The coded signals are presented one pair at a time and in random order to each listener. For each pair of coded signals, listeners have to give their preference for one speech signal. The subjective tests have shown that the average preference score is slightly in favor of the SMVQ technique.

Table 3: Selection of the optimal codebook.

$e_n = \sum_{i=0}^{i=15} r_n^2(i)$
if $e_n \leq \varepsilon$
  $optimal\ codebook = VCB$
else
  $optimal\ codebook = UCB$
end

Table 4: Objective performance of the BMVQ technique.

Speaker	SegSNR (dB)	
	SMVQ	BMVQ
Female	10.65	10.62
Male	9.90	9.74
Average	10.275	10.18

Table 5: Spectral Distortion of the BMVQ technique.

Technique	Avg SD (dB)	Outliers (in %)	
		2-4 dB	> 4 dB
SMVQ	1.31	2.41	0.06
BMVQ	1.32	2.44	0.08

7 CONCLUSIONS

The objective measures illustrate that the performance of the 1.7 kbit/s BMVQ approach are comparable to that of the 2.3 kbit/s SMVQ method. The BMVQ average coding rate is reduced by 0.6 kbit/s comparatively to the coding rate of the G.722.2 SMVQ combination. However, the BMVQ technique is still not robust in the presence of background noise. The efficiency of the BMVQ approach for highly noisy

speech is affected. The correlation between two consecutive LPC vectors is not too high in noisy speech, even for voiced stationary segments. Unlike in clean speech, the number of wrong decisions in the selection of the optimal codebook might increase in noisy voiced segments; an UCB may be selected to quantize the ISF vectors of a voiced frame. Since each of the two BMVQ codebooks is optimized principally for only one class of speech (voiced or unvoiced), a wrong decision in the selection of the optimal codebook will generate relatively high quantization errors.

In future work, we plan on enhancing the robustness of the BMVQ approach in the presence of background noise. We intend designing a prefilter that will reduce (or cancel) the background noise before the speech coding process.

REFERENCES

Y. Agiomyrgiannakis and Y. Stylianou, "Conditional Vector Quantization for Speech Coding", *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 15, n^o 2, pp. 377-386, February 2007.

R. A. Salami, et al., "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder", *IEEE Transactions on Speech and Audio Processing,* vol. 6, n^o 2, pp. 116-130, March 1998.

ITU-T G.722.2, Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB), July 2003.

B. Bessete, et al., "The adaptive multirate wideband speech codec (AMR-WB)", *IEEE Transactions on Speech and Audio Processing,* vol. 10, n^o 8, pp. 620-636, November 2002.

M. Tamni, M. Jelinek, and V. T. Ruoppila, "Signal modification method for variable bit rate wideband speech coding", *IEEE Transactions on Speech and Audio Processing,* vol. 13, n^o 5, pp. 620-636, September 2005.

M. Jelinek and R. Salami, "Wideband speech coding Advances in VMR-WB standard", *IEEE Transactions on Speech and Audio Processing,* vol. 15, n^o 4, pp. 1167-1179, May 2007.

D. Guerchi, T. Rabie, and A. Louzi, "Voicing-based codebook in low-rate wideband CELP coding", *in Proc. of the tenth European Conference on Speech Communication and Technology (Interspeech 2007-Eurospeech),* Antwerp, Belgium, August 2007, pp. 2505-2508.

K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", *IEEE Transactions on Speech, and Audio Processing,* vol. 1, no. 1, January 1993.