Constrained Voicing-Based Codebook in Low-Rate Wideband CELP Coding

Driss Guerchi
College of Information Technology, UAE University, Al Ain, UAE
Email: guerchi@uaeu.ac.ae

***Abstract*— In this work we propose an efficient technique to quantize the immitance spectral frequency (ISF) in an algebraic code-excited linear prediction (ACELP) wideband codec. The Constrained Voicing-Based Vector Quantization (C-VBVQ) presented in this paper improves substantially the performance of the unconstrained-VBVQ approach for both clean and noisy speech. Both techniques reduce the codebook search time by almost one third. However, in the C-VBVQ training phase, the three codebooks that are individually designed for voiced, unvoiced and transition speech, are jointly reoptimized to impose a structural configuration. The proposed technique reduces the processing delay since it restricts the quantization of an input vector to only one smaller but optimal codebook. For each speech frame, one codebook is selected from the set of three codebooks based on the interframe correlation of the spectral information. The C-VBVQ was successfully implemented in an ACELP wideband coder. The objective and subjective performance are not only superior to that of the combination of the split vector quantization and multistage vector quantization but also to the unconstrained VBVQ.**

***Index Terms*— Wideband speech coding, constrained voicing-based vector quantization, low-complexity ISF quantization**

I. INTRODUCTION

Code-excited linear prediction (CELP) coding is the most popular compression approach in many speech-based communication systems. It represents a good trade-off between the main four attributes of a speech coder: coding rate, coded-speech quality, delay and complexity. While in bandwidth-limited applications, the coding rate is of major importance, in some realtime applications the complexity attribute is more important. The computational complexity could generate a processing delay that is not tolerable by these time-sensitive applications. The processing delay which is related to the speech coder complexity may represent a major obstacle to its deployment in some realtime communication systems. Modern realtime applications, such as Voice over IP (VoIP), Internet telephony, video conferencing, video telephony, and wideband telephony, require high-quality speech coders with low processing delay. The ITU-T G.722.2 [1] adaptive multirate wideband (AMR-WB) coder, which meets most of the implementation criteria, is adopted in some time-sensitive applications, such as the Nokia mobile handsets. However, this standard, as well as other wideband codecs, still suffers from high computational complexity when implemented in remote conferencing systems.

In order to minimize the encoding complexity, current code-excited linear prediction coders ignore the interaction of the vocal tract shape (modeled by a set of linear prediction (LP) parameters) with the vocal cords pitch [2]. This suboptimal approximation is reflected in the disjoint operation of spectral quantization and pitch analysis. A major component in CELP coding is the quantization of the LP parameters. In most of the LP model based coders, this quantization is done on a time frame basis without fully exploiting the interframe correlation with past frames [3]. To model the spectral shaping filter with constant coefficients, the speech signal is considered stationary over an analysis window. To achieve a fixed bit rate, the analysis window size is kept constant in most of the speech coding standards, regardless of the acoustical nature of the speech frame.

The optimal search of the best match to an input linear prediction (LP) vector is performed among all codevectors of the codebook. While this method provides the best performance in terms of speech quality, its high complexity prevents its implementation in realtime applications. Several techniques have been proposed to reduce the computational complexity of the LP coefficients (LPC) quantization at the cost of some deterioration in speech quality [4]. The tree search and the multistage vector quantization are widely used to encode the spectral parameters [5]. These quantization approaches speed up the search procedure but increase the memory requirements. In [6] the authors introduce a new channel-optimized multistage vector quantization of LP parameters. Their approach, which reduces the sensitivity to errors in the received VQ indices, consists of jointly designing the stage codebooks using a source and channel-dependant distortion measure. However, the VQ complexity implementation, mentioned in the paper's introduction as a second limitation of VQ, is unexpectedly ignored in further development and discussions.

In this paper we introduce an optimized technique exploiting the interframe spectral information correlation to reduce the search time for the optimal quantized LPC vector. The voicing-based vector quantization (VBVQ), introduced in [7] and substantially improved in this work,

This paper is based on "Voicing-Based Codebook in Low-Rate Wideband CELP Coding," which appeared in the Proceedings of Interspeech 2007, Antwerp, Belgium, August 2007. © 2007 Interspeech.

This work was supported by the Research affairs at the UAE University.

exploits the interframe correlation of the LP parameters to limit the quantization process to a smaller codebook. The LP filter coefficients, as well as their different representations in the frequency domain, show some interframe redundancy. This redundancy, which is easily noticeable in voiced speech, could be used to predict the current LP parameters from those of the past frames.

In the training phase of the constrained voicing-based vector quantization (C-VBVQ) framework, three disjoint ISF (Immitance Spectral Frequency; the LPC representation used in the G.722.2 standard) codebooks are individually populated from voiced, unvoiced, and transition speech. The C-VBVQ differs from the unconstrained VBVQ in the way the codebooks training is performed. In the C-VBVQ approach, the VBVQ codebooks are reoptimized to introduce some structural behavior in their configuration. Since only one codebook is used in the spectrum encoding of each speech frame, the LPC quantization process in the C-VBVQ technique is preceded by the selection of the appropriate codebook. This selection is based on the interframe correlation of the current and previous frame LPC vectors. The C-VBVQ could be considered as a special delayed-decision technique. This type of coding techniques has proven its efficiency in producing very high speech quality [8]. Classification of speech for efficient coding was addressed in many multi-mode coding frameworks. However, our approach focuses on reducing the coder search complexity while producing high coded-speech quality at no extra bit resources. The hybrid coding proposed in [9] combines a frequency-domain parametric coder for voiced and unvoiced speech with a time-domain waveform coder for unvoiced speech. This three mode coder achieves high-quality coded speech at a rate of 4 kb/s but at the expense of high computational complexity.

II. Motivation

A. General-purpose LPC codebooks

Most of the recent speech coder standards use vector quantization (VQ) to code the spectral information. While VQ techniques reduce the coding rate, they increase the search computation load drastically. The performance of VQ schemes increase with codebook size. A codebook with more codewords certainly excels in coding efficiently the spectral parameters. This is achieved to the detriment of an increase in coding rate and computational complexity.

For example, in the G.729 narrowband codec standard [10], a combination of multistage VQ (MVQ) and split VQ (SVQ) is used to determine which 10-dimensional LSF (Line Spectral Frequency, the most popular representation of the LP coefficients in the frequency domain [11]) vector corresponds most closely to the set of LSF input parameters. In the first stage of the search procedure, a codebook of 128 entries is searched; in the second stage two codebooks of 32 entries each are examined, for a total of 192 entries. In the G.722.2 wideband coding standard [12], the same VQ technique, with slight modifications, is employed to code 16 ISF coefficients. A total of 896 (256+256+64+128+128+32+32) entries are tested against the input ISF vector for all the codec modes, except for the 6.60 kbit/s coder which searches the closed codewords among 832 (256+256+128+128+64) entries. These numbers show the degree of the computational complexity of wideband coding, even when using suboptimal VQ techniques such as MVQ and SVQ. The above codebooks could be classified as general-purpose codebooks since they are used to code the spectral parameters in voiced as well as in unvoiced and transient speech. A first step in easing the implementation of low-rate wideband speech coders in real-time communication applications, such as VoIP conferences, lies in minimizing the coder complexity [13]- [14].

An obvious remedy to the above problem consists of reducing the LPC quantization rate, which is related to the number of entries in the VQ codebooks. This solution however deteriorates the coded speech quality. Another alternative is to confine the search of the closest codeword to a smaller number of codebook entries. This latter approach is widely used in the closed-loop pitch lag search, where an open-loop pitch analysis is performed first in order to limit the closed-loop pitch lag search to few lags around the optimal open-loop pitch period. Even though the objective is the same, conceptually the two approaches are different. The issue here is about how to implement this idea without decreasing the codebook bit rate, and by consequence without deteriorating the coded-speech quality. We believe that in CELP coding, some aspects of the LPC quantization are still not fully utilized.

B. Shortcomings of the traditional CELP analysis

In the CELP encoder, linear prediction analysis of order p is performed first on windowed speech frame $x(m)$, $m = 0, \ldots, M-1$, to extract p linear prediction coefficients $(a_1, a_2, \ldots, a_p)$. These coefficients are used to build the LP analysis filter A(z),

$$A(z) = \sum_{i=1}^{p} a_i z^{-i}. \tag{1}$$

When applied to speech signal, this filter removes the short-term redundancy and outputs an LP error signal $e(m)$ with quasi flat spectrum. The LP analysis process is represented by the following equation:

$$E(z) = X(z)A(z). \tag{2}$$

where $X(z)$ and $E(z)$ are the Z-transform of the windowed speech $x(m)$ and the LP error signal $e(m)$, respectively. In voiced speech, the magnitude spectrum of $e(m)$ contains a variety of harmonics which represent the long-term periodicity in this class of speech.

In high-pitched voices, a harmonic may coincide with a formant peak. In such situation, the LP analysis filter will flatten the speech spectral envelope by removing not only the short-term relevant spectral details but also some energy of this harmonic. This phenomenon is illustrated in Figure 1. In this figure, the third harmonic of the speech spectrum coincides with the first formant. As can be seen in the LP error spectrum, the energy of this harmonic is reduced after LP analysis. According to equation 2, some of this energy is contained in the LP analysis filter $A(z)$. We have come to the conclusion that some LP parameters may bear pertinent information about the long-term periodicity in voiced speech. The current commonly-used LP analysis and LP quantization ignore this LPC-pitch correlation since they are both performed on a time frame pitch-independent basis.

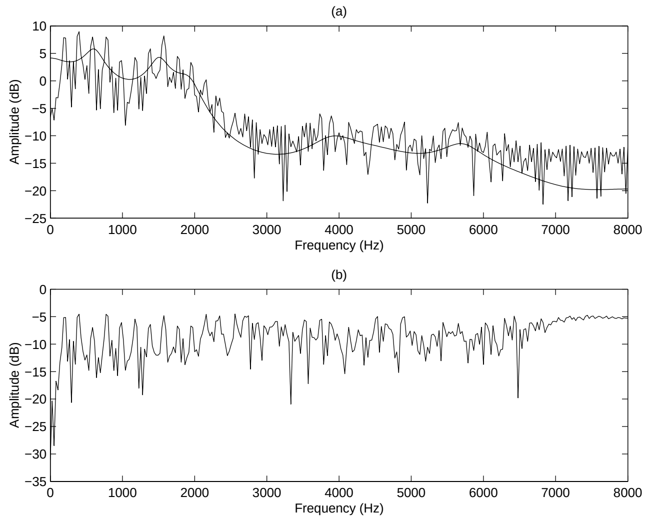

Figure 1. (a) Speech spectrum with the LPC spectrum superimposed; (b) spectrum of the LPC error signal.

To code efficiently the LP filter parameters, one must employ separate codebooks for voiced, unvoiced, and transient speech classes. In both G.729 and G.722.2 standards, the error between the current mean-removed average LSF vector (or ISF in the G.722.2) and the predicted LSF quantized error is coded using a combination of split and multistage vector quantization. In other predictive vector quantization methods, the error between the current and the last quantized LPC vectors (in the LSF representation) is quantized by one or another VQ technique. While the amplitude of this error is too small in voiced speech, its values for unvoiced and transient speech are significantly much higher with a large dynamic range [15]. Figure 2 shows the square of the error $r_{n-1,n}$ between two consecutive ISF vectors, p_{n-1} and p_n in a wideband speech signal. The ISF squared error E_n at frame n is given by:

$$E_n = \sum_{i=1}^{p} (p_n(i) - p_{n-1}(i))^2. \tag{3}$$

III. VOICING-BASED VECTOR QUANTIZATION

In this section, we propose a new source-based encoding technique to reduce the computational complexity of

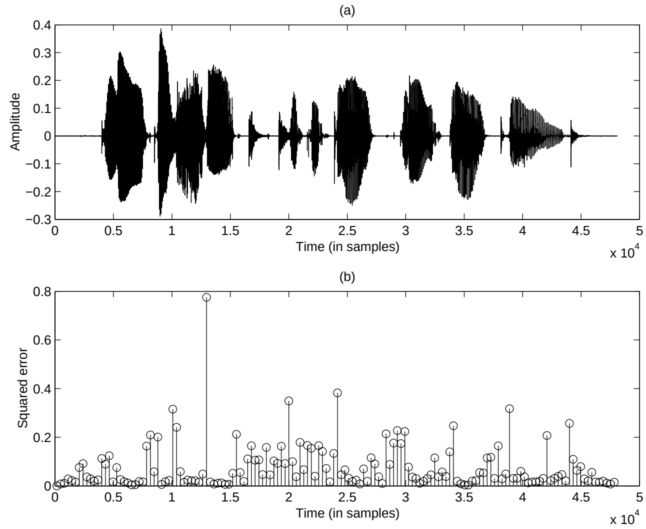

Figure 2. (a) Wideband speech signal; (b) ISF squared error between consecutive frames.

the LPC vector quantization. This technique, called constrained voicing-based multistage vector quantization (C-VBVQ), consists of three disjoint codebooks; a voiced-speech codebook (VCB) quantizes the spectral information in voiced speech, an unvoiced-speech codebook (UCB) codes the unvoiced frames, and a third codebook (TCB) employed in transition speech. This technique looks similar to the split VQ. However, in the split VQ an input vector is divided into two or more subvectors. Each subvector is quantized using a separate codebook. In the C-VBVQ technique, one main virtual codebook is split into three codebooks, but each input vector is quantized using exclusively one codebook. The three codebooks are trained individually from voiced, unvoiced, and transition speech, respectively. The search of the closest match to an input LPC vector is confined to only one codebook. This technique reduces the computational complexity without requiring extra storage or additional bit resources. For each speech frame, LP analysis is performed to extract 16 LP coefficients. The estimated LPC vector is compared (after conversion to an equivalent ISF vector) to the quantized LPC vector of the past frame using a squared error distortion measure. The selection of the optimal codebook for the current frame is based on the relative magnitude of this distortion. We expect that in voiced speech, consecutive LPC vectors are highly correlated. A small error distortion is a cue of quasi-stationary voiced speech.

We propose in this paper to exploit this correlation to estimate the current ISF vector from the past frame ISF coefficients. The motivation behind this idea is shown in Figure 3. The interframe variation of different formants is smooth in voiced segments, unlike unvoiced speech where successive formants show very weak correlation. In the C-VBVQ algorithm, a 1st-order predictor, with coefficient 1, is applied to the input ISF vector. The residual ISF vector r_n is quantized using exclusively either one of the three C-VBVQ codebooks. The details of this algorithm will be given in the next section. Figure 4 illustrates the concept of the C-VBVQ approach.

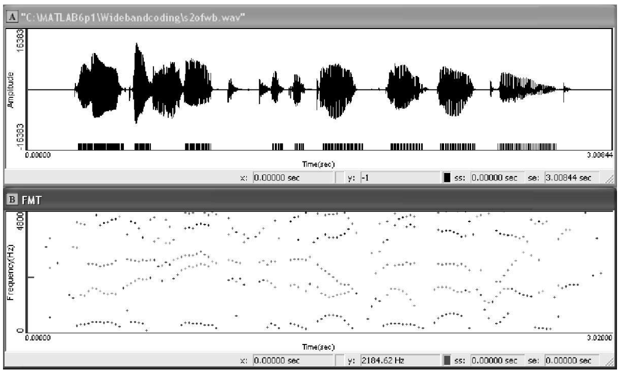

Figure 3. (a) Original wideband speech signal with voiced period marks; (b) Formant trajectory.

IV. Voicing-based codebook design

A. Codebook optimization

In the C-VBVQ paradigm, the three codebooks are trained from a large speech database. In the first phase of the training process, we manipulate manually the speech database to build three sets of speech segments. The first set contains voiced speech, while the second and third are populated from transition and unvoiced speech, respectively. For each frame, we defined the corresponding speech class using three criteria: the signal normalized energy, the zero crossing rate, and the LPC feature with linear discriminant analysis [16]. Since the silence segments have very low energy, the adopted classification criteria tend to join them to the unvoiced speech set. In the second phase, for each set, LP analysis of order 16 is performed every 20 ms using a 30 ms asymmetric window. The resulting LPC vectors are converted to their corresponding ISF vectors. The error between each pair of consecutive ISF vectors is then used in the training of each codebook. Three codebooks, VCB, TCB, and UCB using a combination of SVQ and MVQ, are respectively designed for the three types of speech segments. The procedure is similar to the one used in the G.722.2 standard but it is applied to each of the three codebooks of the C-VBVQ framework. Table I illustrates the algorithm of the C-VBVQ codebooks training. Unlike the unconstrained VBVQ framework, in which the three codebooks are designed independently, the training of the codebooks in the present work is structured. The VCB is reoptimized, according to the UCB design, to guarantee that the energy of each of its codevectors (or ISF errors) is always smaller than those of the UCB ISF errors. The TCB is designed separately since their codevectors show significantly higher energy. This constrained design of the VBVQ codebooks improves the performance of the VBVQ approach since it provides robust criteria for selecting the optimal codebook for ISF quantization. These criteria are discussed in the next section.

TABLE I.
C-VBVQ codebooks training algorithm

1) Classify the speech database into voiced, unvoiced and transition speech.
2) Start with the class of unvoiced speech.
3) Perform LP analysis on every two consecutive frames of the current speech class.
4) Transform the resulting LPC vectors to ISF vectors.
5) Compute the error between each consecutive ISF vectors and stored it in an ISF error database.
6) Calculate the energy of these errors and store them temporarily for later use.
7) Design a SVQ-MVQ based codebook from the ISF error database.
8) Find the minimum ISF error energy, $E_{min}^{(UCB)}$.
9) Find the maximum ISF error energy, $E_{max}^{(UCB)}$ and define it as ϵ_2.
10) Repeat steps 3 to 7 for the class of voiced speech.
11) Find the maximum ISF error energy for voiced speech, $E_{max}^{(VCB)}$.
12) If $E_{max}^{(VCB)} < E_{min}^{(UCB)}$
    go to step 14.
    else
    a) update the class of voiced speech by replacing the speech frames providing errors greater than $E_{min}^{(UCB)}$ by new ones (producing errors smaller than $E_{min}^{(UCB)}$) from the speech database,
    b) go to step 10.
13) Define $E_{max}^{(VCB)}$ as ϵ_1.
14) Repeat steps 3 to 7 for the class of transition speech.

B. Codebook bit allocation

To achieve a fixed bit rate, the three codebooks are allocated the same amount of bits (44 bits). It is worth mentioning that for the VCB codebook, a much smaller bit rate is sufficient to produce competitive objective and subjective performance to that of the G.722.2 codec. The error between two consecutive ISF vectors is too small in voiced speech. Its small variance allows significant bit rate reduction.

At the end of the training process, each codebook will be characterized by one index l, ranging from 0 to 2. Table II shows the bit allocation of the C-VBVQ method. r_n is the ISF error vector between the current ISF vector and the last frame quantized vector. In the first stage, this vector is split into two subvectors $r_{n,1}$ and $r_{n,2}$ of 9 and 7 coefficients, respectively. Then 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 and two subvectors, respectively. A coding rate of 46 (44+2) bits per ISF vector is required to encode the spectral information. The codebook index is encoded as side information with two bits. A total of 640 entries are to be tested over the two stages of the quantization. This is a reduction of almost 30% compared to the G.722.2 MVQ-SVQ. Table III shows the number of the entries per vector to be tested in the C-VBVQ.

V. Selection of the optimal codebook

For every frame, the input ISF vector, p_n, is compared to the last frame quantized ISF vector $\hat{p}_{n-1}$. A comparator

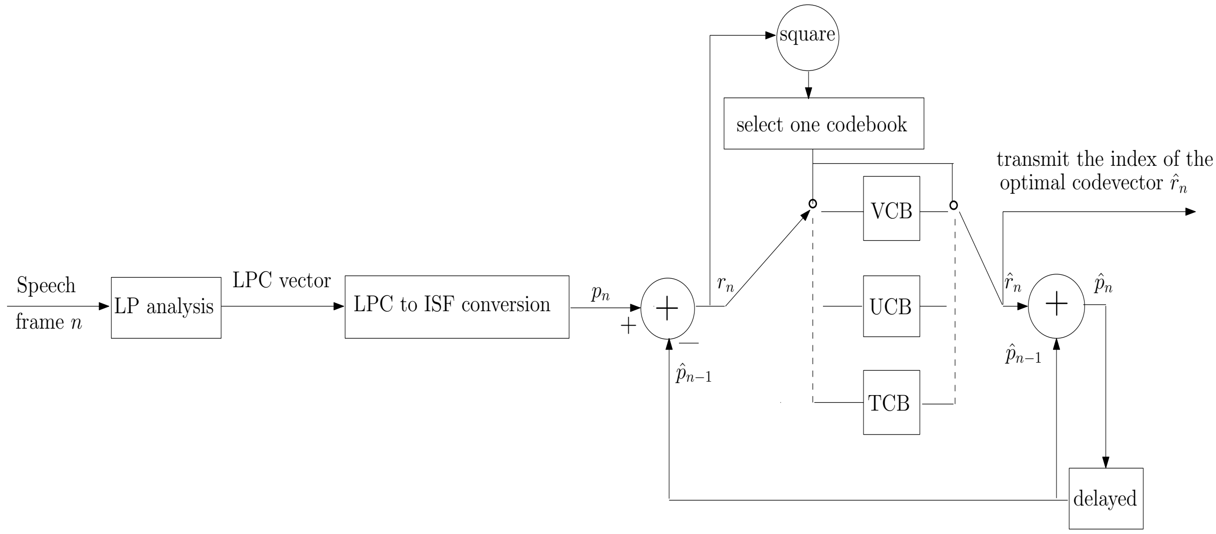

Figure 4. (a) Concept of the constrained voicing-based vector quantization.

TABLE II.
BIT ALLOCATION OF THE C-VBVQ

Stage 1	stage 2
r_1 (7 bits)	$(r-\hat{r}_1)(0:2)$ (6 bits) $(r-\hat{r}_1)(3:5)$ (7 bits) $(r-\hat{r}_1)(6:8)$ (7 bits)
r_2 (7 bits)	$(r-\hat{r}_2)(0:2)$ (5 bits) $(r-\hat{r}_2)(3:6)$ (5 bits)

TABLE III.
NUMBER OF TESTED ENTRIES PER ISF ERROR IN THE C-VBVQ

Input vector	Number of entries
r_1	128
r_2	128
$(r-\hat{r}_1)(0:2)$	64
$(r-\hat{r}_1)(3:5)$	128
$(r-\hat{r}_1)(6:8)$	128
$(r-\hat{r}_2)(0:2)$	32
$(r-\hat{r}_2)(3:6)$	32
Total	640

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_1 (defined in Table I as $E_{max}^{(VCB)}$), the VCB will be used for the search of the closest vector to the input vector p_n. Otherwise UCB or TCB will be adopted according to the value of the square of r_n relative to another threshold ϵ_2, which is introduced in Table I as $E_{max}^{(UCB)}$. The advantage of this method is that for steady-state speech frames, the percentage of hitting the optimum in the VCB codebook is greater than 95%. For non-stationary speech, such as unvoiced and transition between different phonemes, the UCB and the TCB will be used, respectively. Table IV illustrates the algorithm for optimal codebook selection.

TABLE IV.
SELECTION OF THE OPTIMAL CODEBOOK

$E_n = \sum_{i=0}^{i=15} r_n^2(i)$
$if\ E_n \leq \epsilon_1$
  $optimal\ codebook = VCB$
$elseif\ \epsilon_1 < E_n \leq \epsilon_2$
  $optimal\ codebook = UCB$
$elseif\ \epsilon_2 < E_n$
  $optimal\ codebook = TCB$
end

VI. EVALUATION

We have conducted several simulations to compare the performance, in terms of objective and subjective measures, of the C-VBVQ to the common SVQ-MVQ combination approach and unconstrained VBVQ. The codebooks in the three techniques have been trained using the same database. This is to avoid any effects of the type of the database on the performance comparison. As an objective quality measure, we selected the segmental signal-to-noise Ratio (SEGSNR) at the output of the decoder. The SEGSNR represents the average of the SNR over the whole duration of the speech file. For a given speech frame, the SNR is defined as

$$SNR(dB) = 10\log\left(\frac{\sum_{m=0}^{319} s(m)^2}{\sum_{m=0}^{319}[s(m)-\hat{s}(m)]^2}\right). \qquad (4)$$

where $s(m)$ and $\hat{s}(m)$ are the original and reconstructed speech, respectively. The systems to be evaluated are three versions of the same wideband algebraic CELP (ACELP) coder. The three coders are similar except in the ISF quantization, where in the first coder we use a standard combination of SVQ-MVQ . In the second and third ACELP coders, we implement the unconstrained VBVQ (U-VBVQ) and C-VBVQ techniques, respectively. To separate the effects of the quantization of the other coder parameters (such as pitch lag and gain, LP excitation signal, and fixed codebook contribution) on the result of the three coders performance, only the spectral information is quantized in the above coders. The database for training phase consists of 150 minutes of English speech from 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 [17] is used to evaluate the ISF quantization in the three versions of the ACELP coder. The evaluation simulations have been conducted on eight different input sentences uttered by other speakers. Table V presents the SEGSNR of the three ACELP versions for clean speech. Table VI shows the average spectral distortion between the input ISF vectors and their corresponding quantized ISF vectors after implementing the three quantization techniques for clean speech. The improvements of the C-VBVQ over the U-VBVQ approach are shown in Table VII and Table VIII for noisy speech.

Informal listening tests, represented by the comparative mean opinion score (CMOS), has been carried out as subjective measure tool. The C-VBVQ coded speech is compared separately with the SVQ-MVQ and U-VBVQ coded speech. We rated, on a 3-point scale (-1, 0, 1), the listeners opinion on the better quality among the C-VBVQ versus U-VBVQ and SVQ-MVQ coded speech. Point 1 is marked if a listener chooses the C-VBVQ coded speech, -1 if the rival technique (SVQ-MVQ or U-VBVQ) , and 0 if a listener couldn't report any clear difference between both signals. Table IX and Table X depicts the high subjective performance of the C-VBVQ approach with regard to both the SVQ-MVQ and U-VBVQ.

TABLE V.
SEGSNR COMPARISON OF SVQ-MVQ, U-VBVQ, AND C-VBVQ FOR CLEAN SPEECH

Speaker	SEGSNR (dB)		
	SVQ-MVQ	U-VBVQ	C-VBVQ
Female	10.72	11.36	11.41
Male	9.88	10.55	10.60
Average	10.3	10.955	11.005

TABLE VI.
SPECTRAL DISTORTION COMPARISON OF COMMON SVQ-MVQ, U-VBVQ, AND C-VBVQ FOR CLEAN SPEECH

ISF Quantization	Avg SD (dB)	Outliers (in %)	
		2-4 dB	$>$ 4 dB
SVQ-MVQ	1.29	2.34	0.04
U-VBVQ	1.20	2.36	0.07
C-VBVQ	0.92	2.01	0.03

TABLE VII.
SEGSNR COMPARISON OF COMMON SVQ-MVQ, U-VBVQ AND C-VBVQ FOR NOISY SPEECH

Speaker	SEGSNR (dB)		
	SVQ-MVQ	U-VBVQ	C-VBVQ
Female	10.10	10.90	11.38
Male	9.70	10.14	10.50
Average	9.9	10.52	10.94

VII. DISCUSSION AND CONCLUSIONS

The objective measure illustrates that the performance of the C-VBVQ approach are slightly superior to the standard combination of SVQ-MVQ whilst the complexity is reduced by almost one third in the C-VBVQ. The informal listening tests reveal that the three techniques provide competitive overall subjective quality for both clean and noisy speech. For voiced clean speech, the U-VBVQ and C-VBVQ subjective performance are significantly higher than that of the SVQ-MVQ. However, the C-VBVQ produces better coded-speech quality for voiced noisy speech. The C-VBVQ technique, as well as the U-VBVQ, reduces the execution time of the ISF parameters quantization when operating at the same bit rate of the G.722.2 SVQ-MVQ method. The overall complexity expense of the spectral parameters coding is reduced by almost 30 %. The major drawback of the U-VBVQ is that its efficiency for noisy speech is negatively 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 increases in noisy voiced segments; UCB might be selected to quantize the ISF vectors of a voiced frame. Since each of the three codebooks of the U-VBVQ technique is optimized specifically for only one type of speech (voiced, unvoiced or transition), a wrong decision in the selection of the optimal codebook will generate high quantization errors. The C-VBVQ accommodates this drawback since it imposes some robust criteria for selecting the best codebook. The structural configuration of the C-VBVQ codebooks, leading to three disjoint optimal codebooks, is granted by the design constraints on the training process. Even though not explicitly designed for noisy channels, the C-VBVQ is fairly robust to propagation errors in the transmitted ISF codevector index. When a transmission error occurs, the codevector at the decoder is still obtained from the same selected codebook at the encoder. This will limit the quantization error to an upper worst-case value. For example, when the transmission error occurs in the VCB optimal ISF index, the quantization error is still smaller than ϵ_1 since at the decoder another vector from the same VCB will be used instead. In the common SVQ-MVQ, G.722.2-adopted technique, a channel error in the ISF vector may lead to the use of a different-speech class ISF vector at the decoder.

ACKNOWLEDGMENT

The authors would like to thank Research Affairs at the UAE University for funding.

TABLE VIII.
SPECTRAL DISTORTION COMPARISON OF COMMON SVQ-MVQ, U-VBVQ AND C-VBVQ FOR NOISY SPEECH.

ISF Quantization	Avg SD (dB)	Outliers (in %)	
		2-4 dB	> 4 dB
SVQ-MVQ	1.32	2.42	0.07
U-VBVQ	1.26	2.46	0.09
C-VBVQ	0.95	2.07	0.05

TABLE IX.
COMPARISON MEAN OPINION SCORE OF THE U-VBVQ TECHNIQUE FOR CLEAN SPEECH.

Speaker	CMOS	
	C-VBVQ vs SVQ-MVQ	C-VBVQ vs U-VBVQ
Female	0.94	0.86
Male	0.90	0.82
Average	0.92	0.84

TABLE X.
COMPARISON MEAN OPINION SCORE OF THE U-VBVQ TECHNIQUE FOR NOISY SPEECH.

Speaker	CMOS	
	C-VBVQ vs SVQ-MVQ	C-VBVQ vs U-VBVQ
Female	0.86	0.91
Male	0.83	0.82
Average	0.845	0.865

REFERENCES

[1] Varga, I., Iacovo, R. D. and Usai, P., "Standardization of the AMR wideband speech codec in 3GPP and ITU-T", *IEEE Communication Magazine,* vol. 44, n^o 5, pp. 66-73, May 2006.

[2] D. Guerchi, Y. Qian, and P. Mermelstein, "Pitch-synchronous linear-prediction analysis by synthesis with reduced pulse densities", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, June 2000, vol. 3, pp. 1491-1494. .

[3] D. Guerchi and P. Mermelstein, "Low-rate quantization of spectral information in a 4 kb/s pitch-synchronous CELP coder", *IEEE Workshop on Speech Coding*, Delavan, WI, September 2000, pp. 111-113.

[4] Kang, S., Shin, Y. and Fischer, T. R., "Low-complexity predictive trellis coded quantization of speech line spectral frequencies", *IEEE Transactions on Signal Processing,* vol. 52, pp. 2070-2079, July 2004.

[5] A. Gersho and R. M. Gray, "Vector quantization and signal compression". *Norwell, M.A., Kluwer*, 1992.

[6] V. Krishnan, D.V. Anderson, and K. K. Truong, "Optimal multistage vector quantization of LPC paramters over noisy channels", *IEEE Transactions on Speech and Audio Processing,* vol. 12, n^o 1, pp. 1-8, January 2004.

[7] D. Guerchi, T. Rabie, and A. Louzi, "Voicing-based codebook in low-rate wideband CELP coding", *Proceedings of Interspeech (Eurospeech) 2007,* Antwerp, Belgium, August 2007, pp.2505-2507.

[8] K. Mano and T. Moriya, "4.8 kbit/s delayed decision CELP coder using tree coding," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1990, pp. 21-24.

[9] Shlomot, E., Cuperman, V. and Gersho A., "Hybrid coding: Combined harmonic and waveform coding of speech at 4 kb/s", *IEEE Transactions on Speech and Audio Processing,* vol. 9, n^o 6, pp. 632-646, September 2001.

[10] Salami, R.A. 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.

[11] F. Itakura, "Line spectrum representation of linear predictive coefficients," *Journal of Acoustics Society of America,* vol. 57, no. 1, pp. S35, 1975.

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

[13] Hiwasaki, Y. et al., "A G.711 Embedded Wideband Speech Coding for VoIP Conferences", *IEICE Transactions on Information and Systems,* vol. E89-D, n^o 9, pp. 2542-2551, September 2006.

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

[15] T. Eriksson and F. Norden, "Memory-based vector quantization of LSF parameters by power series approximation", *IEEE Transactions on Speech and Audio Processing,* vol. 15, n^o 4, pp. 1146-1155, January 2004.

[16] J.K. Shah, A.N. Iyer, B. Y. Smolenski, and R. E. Yantorno, "Robust voiced/unvoiced classification using novel features and gaussian mixture model," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 17-21, Montreal, Canada, 2004.

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

Driss Guerchi is currently an Assistant Professor at the UAE University, UAE. He received his PhD in Telecommunications from the National Institute for Scientific Research, Montreal, Canada in 2001, and his Master degree in Physics from University of Quebec at Montreal, Montreal, Canada in 1994.

He was a lecturer at the National Institute for Scientific Research Montreal, 2000-2001, and at the University of Quebec at Montreal, Montreal, Canada, 1994-1995. His research interests include speech coding, speech steganography and digital signal processing.