Computational Reduction For G.729's LSP Quantization

Cheng-Yu Yeh
Department of Electrical Engineering, National Chin-Yi University of Technology, Taichung, Taiwan
cy.yeh@ncut.edu.tw

Chang-Zhi Zhuo
Department of Electrical Engineering, National Chin-Yi University of Technology, Taichung, Taiwan
fzr1216@yahoo.com.tw

Abstract

G.729 standard has been widely used in the VoIP system. But the computational complexity is too large to real-time work via software implementation on embedded devices. Even though the G.729A, it is still hard to implement. This paper will focus on the computational reduction of the quantization procedure of LSP coefficients. Thus, the hybrid two-stage VQ is proposed to replace the original structure. Experimental results confirm that more than 80% of computations can be eliminated in comparison with the conventional G.729 as well as the performance of speech quality is still good. The proposed approach can be successfully used in LSP quantization procedure and the efficiency is excellent.

Keywords: CS-ACELP, LSP, vector quantization, tree VQ.

1. Introduction

G.729 is a standard for speech compression publicized by ITUT in 1996 [1]-[6], and this standard of speech compression is mainly to introduce an algorithm for the coding of speech signals using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP). G.729 operates to take every 10ms of the speech frame as an unit with a sampling frequency of 8KHz at a bit rate of 8 bit/s that is applicable to the field of recording pen, VoIP, etc. The CS-ACELP based algorithm and analysis-by-synthesis principle lead to provide good synthesized speech quality at low bit-rate [1]. But the large computational complexity is also induced.

The coding process of G.729 is mainly divided into six blocks; wherein, the input signal is high-pass filtered and scaled in the pre-processing block, and then followed an analysis of Linear Prediction to obtain the LPC parameter. In virtue of that the line-spectrum parameter is characterized with a better interpolation and quantization, LPC parameter is then converted into LSP parameter and quantized by using predictive two-stage quantization. For the sake of homology to the hearing effect of human ears, a signal processing will be made prior to the coding by going through a perceptual weighting filter. Open-loop pitch analysis is principally to reduce the pitch for searching the adaptive codebook that takes every 10ms as the unit for computation. After roughly determining the range of pitch, a further precise pitch can be located when searching in the closed loop; the major design of the fixed codebook search thereafter is originated from the structure of algebraic codebook while each codebook contains four non-zero pulses with either the amplitudes +1 or -1 to each pulse so that the fixed codebook can be searched with the mean-squared error minimized. Figure 1 and Figure 2 show the block diagrams for G.729 encoding process and decoding process respectively.

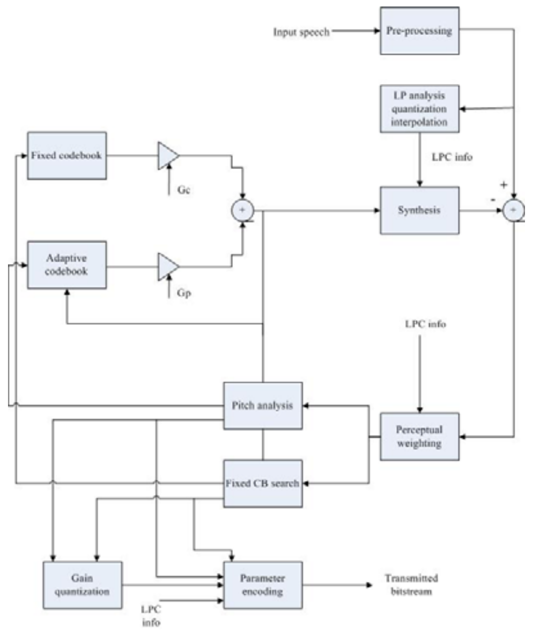

Figure 1. Block diagram of G.729 encoding phase

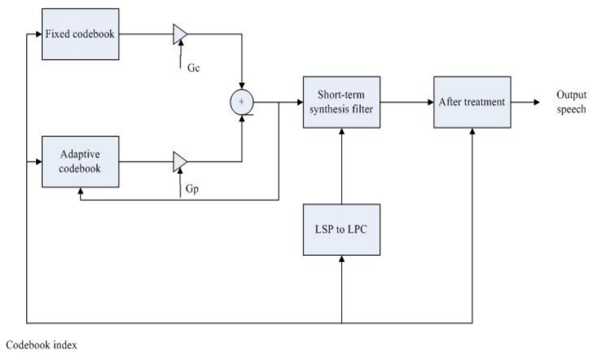

Figure 2. Block diagram of G.729 decoding phase

In the foregoing blocks, the computational load for linear prediction and quantization has occupied

19.49% of the total blocks. Since the distance will be computed by the input vector and codebook during the quantizing process, the smaller distance will be the index while the indices will form an index list. During the subject search process, an iterative computation for the distance will be going on and on that will consume a hefty load of computation.

In this paper, the hybrid two-stage VQ, which applied to the quantization procedure of LSP coefficients, is proposed to replace the original structure for reducing the computational complexity. All of the theoretic analyses and experimental results will be presented to verify the performance of proposed approach.

In the following sections, the linear prediction analysis and quantization is introduced in Section 2. The proposed approach for the application of quantization to reduce the complexity from computational load is depicted in Section 3. Section 4 will show the experimental results and discussions. The conclusion will be given in the last section.

2. Linear Prediction Analysis and Quantization

2.1 Linear prediction analysis

The linear prediction technique, taking the advantage of P^{th}-order linear prediction filter, is the most frequent used technique for speech analysis. In G.729, the P^{th}-order is set to 10^{th}-order and the $\hat{a}_i$ is set as the linear prediction coefficients. The filter is defined as:

$$\frac{1}{\hat{A}(Z)} = \frac{1}{1+\sum_{i=1}^{10} \hat{a}_i Z^{-i}} \qquad (1)$$

The analysis is performed against the input signal every 10ms speech frame through the manipulation of Levison-Durbin algorithm to compute the coefficients of linear prediction filter. In the mean time, the windowing should be made prior to the analysis while the window consists of a Hamming window and a quarter of a cosine signal that can be expressed by Equation (2).

$$w_{lp}(n) = \begin{cases} 0.54 - 0.46\ \cos\left(\frac{2\pi n}{399}\right) & n = 0,\ldots,199 \\ \cos\left(\frac{2\pi(n-200)}{159}\right) & n = 200,\ldots,239 \end{cases} \qquad (2)$$

Every subframe is 5ms, and the windowing action is to window 10ms from the present speech frames, 15ms from the past speech frames and 5ms from the future frames; so there are 30ms and 240 sampling points in total. The windowing procedure is illustrated in Figure 3.

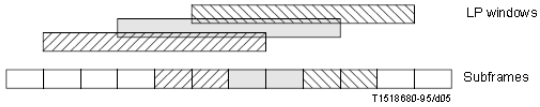

Figure 3. Windowing procedure in LP analysis

2.2 LPC to LSP conversion

The parameter of linear prediction coefficients (LPC) are converted to the linear spectral pair (LSP) coefficients for quantization and interpolation purposes [1]-[2]. For a 10^{th} order LP filter, the LSP coefficients are defined as the roots of the sum and difference polynomials:

$$\begin{aligned} F_1'(z) &= A(z) + z^{-11}A(z^{-1}) \\ F_2'(z) &= A(z) - z^{-11}A(z^{-1}) \end{aligned} \qquad (3)$$

These two polynomial equations are symmetric with each other, all the roots of $F_1'(z)$ and $F_2'(z)$ are on a unit circle and they alternate with each other; that is, $F_1'(z)$ has a root of z=-1 at $\omega = \pi$ while $F_2'(z)$ has a root of z=1 at $\omega = 0$. Two new polynomial equations can be derived by eliminating these two roots as expressed by Equation (4); also, each polynomial has five conjugate roots on the unit circle.

$$\begin{aligned} F_1(z) &= F_1'(z)/(1+z^{-1}) = \prod_{i=1,3,\ldots,9}(1-2q_i z^{-1}+z^{-2}) \\ F_2(z) &= F_2'(z)/(1-z^{-1}) = \prod_{i=2,4,\ldots,10}(1-2q_i z^{-1}+z^{-2}) \end{aligned} \qquad (4)$$

2.3 Quantization of LSP coefficients

The LSP coefficients q_i are quantized using the linear spectral frequencies (LSF) representation ω_i , and the equation of conversion is defined as:

$$\omega_i = \arccos(q_i) \quad i = 1,\ldots,10 \qquad (5)$$

A switched 4^{th}-order MA predictor is used to predict the LSF coefficients of the current speech frame while a differential computation is made between the computed and the predicted coefficients as expressed by Equation (6).

$$l_i = [\omega_i^{(m)} - \sum_{k=1}^{4} \hat{p}_{i,k}\, \hat{l}_i^{(m-k)}]/(1-\sum_{k=1}^{4}\hat{p}_{i,k}) \qquad (6)$$

Where $\hat{p}_{i,k}$ are the coefficients of the switched MA predictor. l_i are the predicted outputs, $\hat{l}_i^{(m-k)}$ are the quantized outputs of l_i from previous several speech frames $(m-k)$; through the two-stage vector quantizer which diagram is shown in Figure 4.

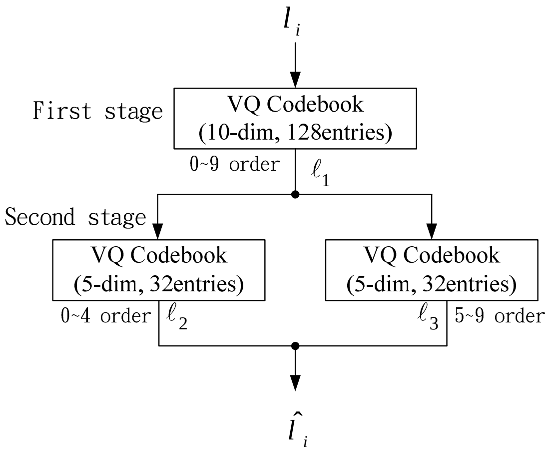

Figure 4. Schematic diagram of two-stage VQ

For the two-stage VQ, the first stage is a 10-dimensinal VQ using ℓ_1 as the index for this codebook with 128 entries (7 bits). The second stage is a split VQ using two 5-dimensional codebooks which containing 32 entries (5 bits) each. Wherein, ℓ_2 and ℓ_3 as the indices for these codebooks respectively. ℓ_2 represents the lower part of 5 dimensions (0~4) while ℓ_3 represents the higher part of 5 dimensions (5~9). This structure of two-stage VQ can be expressed by Equation (7).

$$\hat{l}_i = \begin{cases} \mathrm{L1}_i(\ell_1) + \mathrm{L2}_i(\ell_2) & i = 1,\ldots,5 \\ \mathrm{L1}_i(\ell_1) + \mathrm{L3}_{i-5}(\ell_3) & i = 6,..,10 \end{cases} \tag{7}$$

After the quantization, the quantized LSF parameters $\hat{\omega}_i^{(m)}$ for the current frame (m) can be obtained as follows:

$$\hat{\omega}_i^{(m)} = (1 - \sum_{k=1}^{4} \hat{p}_{i,k})\,\hat{l}_i^{(m)} + \sum_{k=1}^{4} \hat{p}_{i,k}\ \hat{l}_i^{(m-k)} \quad i = 1,\ldots,10 \tag{8}$$

The best LSF parameters can be obtained by taking the approximation defined by Equation (9) that minimizes the weighted mean-square error from two MA predictors.

$$E_{lsf} = \sum_{i=1}^{10} w_i(\omega_i - \hat{\omega}_i)^2 \tag{9}$$

The weights w_i are made adaptive as a function of the unquantized LSF coefficients defined by Equation (10); wherein, the weights w_5 and w_6 are multiplied by 1.2 each.

$$w_i = \begin{cases} 1.0 & if\ \omega_2 - 0.04\pi - 1 > 0 \\ 10(\omega_2 - 0.04\pi - 1)^2 + 1 & otherwise \end{cases} \tag{10}$$

$$w_i\ 2 \le i \le 9 = \begin{cases} 1.0 & if\ \omega_{i+1} - \omega_{i-1} - 1 > 0 \\ 10(\omega_{i+1} - \omega_{i-1} - 1)^2 + 1 & otherwise \end{cases}$$

$$w_{10} = \begin{cases} 1.0 & if\ -\omega_9 + 0.92\pi - 1 > 0 \\ 10(-\omega_9 + 0.92\pi - 1)^2 + 1 & otherwise \end{cases}$$

3. The Proposed Algorithm

The hybrid two-stage VQ is proposed and shown in Figure 5 to reduce the complexity. In this structure, the first stage and the second stage are modified from general VQ to binary tree VQ.

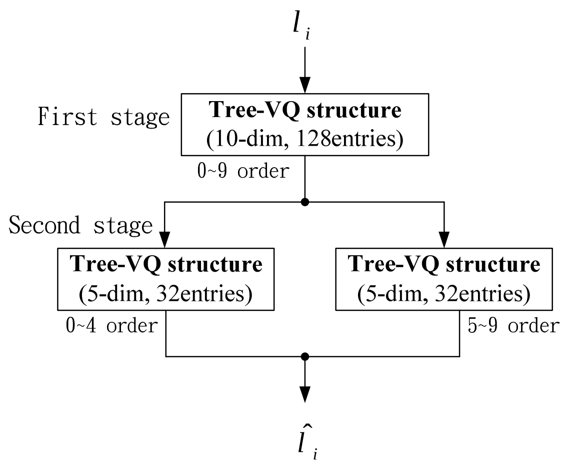

Figure 5. Schematic diagram of hybrid two-stage VQ

Before using the binary tree VQ, the 128 entries of codebook in the first stage must be trained firstly to construct the balanced tree. This procedure is performed in the off-line training. Thus, seeking the best codeword can be achieved by passing through 7 nodes and 14 entries only. Also, the training procedure of tree VQ in the second stage is the same as the first stage. In contrast with the original method, about 89% of complexity can be discarded in the first stage.

4. Experimental Results

In order to examine the efficiency of complexity reduction for the proposed methodology, the experiments with computing saving (CS) are presented. Besides, the perceptual evaluations of speech quality; PESQ-MOS and MOS-LQ [7]-[9], are employed and calculated to inspect the performance of proposed algorithm.

A PC with 2.6G Hz CPU and Windows XP operating system was used as the tool for experiment, and the program was written in Visual C++ 6. Moreover, the testing data consists of two speech databases, 655 sentences of male speech and 655 sentences of female speech respectively, as shown in Table 1.

Table 1. Information of testing data

Database	Speech frame count	Syllable count	File count
Male speech	971616	35243	655
Female speech	1161536	35244	655

Table 2 and Table 3 show that the computing savings on male/female testing data. The average time of G.729 and proposed approach were obtained by calculating the average value from ten repeatedly performed experiments. Furthermore, the computing savings on male/female can be obtained too.

In Table 2, it is evident to show that about 82.9% of computations can be eliminated just applying the tree VQ in the first stage only. This efficiency is very outstanding. Also, totally 83.18% of CS can be achieved in the whole structure applying the tree VQ.

In Table 3, totally 79.17% of CS can be obtained in the whole structure applying the tree VQ on female testing data. These results are similar to the experiment on male testing. Both experimental results confirm that the efficiency of proposed approach is outstanding.

Table 2. The CS comparison for male testing data

	Stage 1 only	Entire Procedure
Average Time of G.729 (ms/frame)	0.0193	0.0214
Average Time of Proposed (ms/frame)	0.0033	0.0036
Computing Savings (%)	0.8290	0.8318

Table 3. The CS comparison for female testing data

	Stage 1 only	Entire Procedure
Average Time of G.729 (ms/frame)	0.0177	0.0216
Average Time of Proposed (ms/frame)	0.0034	0.0045
Computing Savings (%)	0.8079	0.7917

On the other hand, the perceptual evaluation of speech quality; PESQ-MOS and MOS-LQ are calculated to inspect the performance of speech quality. The range of PESQ-MOS score is between 1.0 and 4.5. The experiments in Table 4 show that the PESQ-MOS score of the G.729 and the proposed approach is 3.771 and 3.679 respectively. Furthermore, Table 5 shows that the PESQ-MOS score of the G.729 and the proposed approach is 3.293 and 3.216 respectively. These results confirm that the performance degradation of proposed approach is slightly, the speech quality is still good.

Table 4. The speech quality comparison for male

Method	PESQ-MOS	MOS-LQ
G.729	3.771	3.901
Proposed	3.679	3.787

Table 5. The speech quality comparison for female

Method	PESQ-MOS	MOS-LQ
G.729	3.293	3.255
Proposed	3.216	3.144

5. Conclusions

In this paper, we have performed the experiments by applying the hybrid two-stage VQ approach to reduce the computational load from the conventional VQ. According to the experimental results, it is shown that this algorithm can reduce the computational load up to 80% from the testing data of both male speech and female speech. In the meantime, the performance degradation of proposed approach is slightly, the speech quality is still good. These results confirm that the proposed approach can be successfully used in LSP quantization procedure and the efficiency is excellent.

References

[1] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using CS-ACELP", March 1996.

[2] R. Salami, et al., "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder", IEEE Trans. Speech and Audio Processing, Vol. 6, No. 2, pp. 116-130, March 1998.

[3] R. Salami, et al., "ITU-T G.729 AnnexA: Reduced Complexity 8kb/s CS-ACELP Codec for Digital Simultaneous Voice and Data", IEEE Communication Magazine, pp. 56-63, Sept. 1997.

[4] Perkins M. E., et al., "Characterizing the subjective performance of the ITU-T 8kb/s speech coding algorithm-ITU-T G.729", IEEE Communications Magazine, Vol. 35, No. 9, pp. 74-81, Sept. 1997.

[5] S. H. Hwang, "Computational improvement for G.729 standard," *Electronics Letters*, Vol. 36, No. 13, pp. 1163-1164, June 2000.

[6] C. Laflamme, J-P. Adoul, R. Salami, S. Morissette, and P. Mabilleau, "16 kbps wideband speech coding technique based on algebraic CELP," in *Proc. ICASSP*, Vol. 1, pp. 13-16, April 1991.

[7] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Feb. 2001.

[8] ITU-T Recommendation P.862 Amendment 1, "Source code for reference implementation and conformance tests," March 2003.

[9] A. E. Conway, "Output-based method of applying PESQ to measure the perceptual quality of framed speech signals", in *IEEE Wireless Communications and Networking Conference*, Vol. 4, pp. 2521-2526, March 2004.