International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:6, No:10, 2012

Enhanced Spectral Envelope Coding based on NLMS for G.729.1

Keunseok Cho, Sangbae Jeong, Hyungwook Chang, and Minsoo Hahn

Abstract**—**In this paper, a new encoding algorithm of spectral envelope based on NLMS in G.729.1 for VoIP is proposed. In the TDAC part of G.729.1, the spectral envelope and MDCT coefficients extracted in the weighted CELP coding error (lower-band) and the higher-band input signal are encoded. In order to reduce allocation bits for spectral envelope coding, a new quantization algorithm based on NLMS is proposed. Also, reduced bits are used to enhance sound quality. The performance of the proposed algorithm is evaluated by sound quality and bit reduction rates in clean and frame loss conditions.

Keywords**—**G.729.1, MDCT coefficient, NLMS, spectral envelope.

I. INTRODUCTION

RECENTLY, the technology of embedded speech and audio coding algorithms is rapidly advancing. The enhancement of coding algorithms is an interesting issue and is being studied by many research organizations. G.729.1 is a new scalable speech and audio coder, which is a scalable extension of the ITU-T G.729 speech coding standard widely used in voice over IP (VoIP) applications. G.729.1 has been recently standardized by ITU-T for wideband telephony and VoIP applications. During the transmission of a bit stream, the scalability of G.729.1 allows a flexible bit rate and bandwidth adjustment, which can be implemented in gateways or other device combining multiple data streams. G.729.1 can provide smooth migration from narrow band (300-3400Hz) telephony to high quality wideband (50-7000Hz) telephony [1]. G.729.1 can encode wideband signals at the bitrates of 14-32 kbps. An interesting feature of the coder is in the embedded bitstream which has stack structures of 2 kbps allowing good quality improvements [2]. By using the modified discrete cosine transform (MDCT), the weighted code-excited linear prediction (CELP) difference signal in the lower band and the direct higher-band signal are transformed into a frequency domain.

To encode spectral envelope in G.729.1, a lot of bits are needed. To improve the sound quality of G.729.1, we propose a new algorithm for encoding spectral envelope information by normalized least mean square (NLMS) algorithm using past frame sub-band energies in a time-domain alias cancellation (TDAC). To solve the coding error propagation phenomenon in frame error conditions during transmission, the original envelope coding algorithm is used at a given interval. To evaluate the performance of the proposed spectral envelope coding scheme, bit reduction and spectral distortion were measured. Additionally, preference tests were performed.

The organization of the paper is as follows. The TDAC encoder part to apply the proposed algorithm is described in Section II. Section III describes the proposed algorithm in detail. The experimental results of proposed algorithm are summarized and discussed in Section IV. Finally, conclusions are given in section V.

II. TDAC ENCODER OF G.729.1 CODER

The G.729.1 coder is an 8-32 kbps scalable wideband coder by extension of the G.729. The encoder of the G.729.1 is composed of three structures: embedded code-excited linear prediction (CELP) coding of the lower band (50-4000Hz), time-domain bandwidth extension (TDBWE) for parametric coding of the higher band (4000-7000Hz), and TDAC for the enhancement of the full band (50-7000Hz) by predictive transform coding. [2]. The TDAC part jointly encodes the weighted lower-band CELP coding error (residual) and the higher-band signal by the MDCT. The TDAC encoder is shown in Fig. 1. It encodes the full-band signal in MDCT domain at the highest bit rate (32 kbps). The full band signal is composed of the perceptually weighted CELP coding error in the lower band and the raw input signal in the higher band. The MDCTs of the lower and higher band signal are produced and quantized in the TDAC encoder [1]. During the quantization, the joint spectrum of the full band is divided into sub-bands. For each sub-band, the spectral envelop is encoded by uniform quantization along with the Huffman coding scheme, and the shape is encoded by spherical vector quantization.

The spectral envelope is represented by the root mean square (RMS) in log domain as shown in (1).

$$\log\_rms(j) = \frac{1}{2}\log_2\left[\frac{1}{N(j)}\sum_{k=b(j)}^{b(j+1)} Y^2(k) + \varepsilon_{rms}\right], j = 0,...,17 \quad (1)$$

where $\varepsilon_{rms} = 2^{-24}$ and j, k are the sub-band and the MDCT index, respectively. The $N(j)$ and the $b(j)$ are respectively the number of MDCT coefficient and sub-band boundary in j^{th} sub-band. $Y(k)$ is the k^{th} MDCT coefficient of a full band spectrum. The spectral envelope is quantized with 5 bits by uniform scalar quantization. The RMS indices are quantized with 3 dB steps and encoded by the two-mode binary encoder comprising the

Keunseok Cho and Minsoo Hahn are with the Electrical Engineering Department, Korea Advanced Institute of Science and Technology, Daejeon, Korea.

Sangbae Jeong and Hyungwook Chang are with the Electronics Engineering Department/ERI, Gyeongsang National University, Jinju, Korea (Corresponding author to provide phone: +82-55-772-1727,e-mail:jeongsb@gnu.ac.kr).

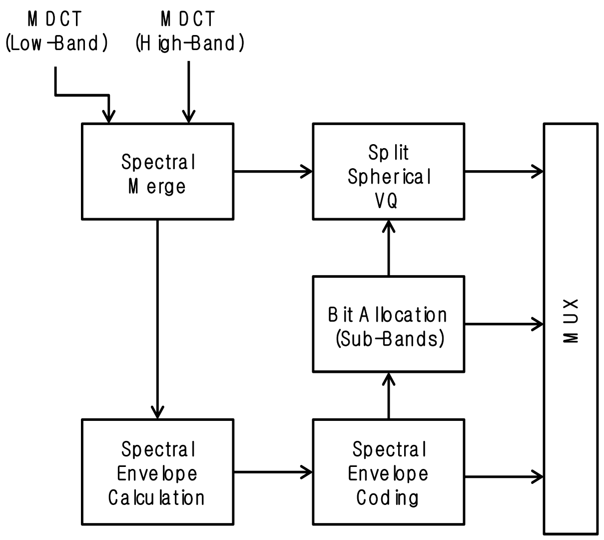

Fig. 1 Block diagram of the TDAC encoder

differential Huffman encoder and the natural binary encoder. The RMS indices are limited between -11 and +20 (32 possible values). The quantized full-band envelope information is split into two sub-vectors that are lower-band spectral envelope (r(0) ~ r(9)) and higher-band spectral envelope (r(10) ~ r(17)). Using two mode lossless coding algorithm, two sub-vectors are coded separately. In order to minimize the average bit transmission, the differential Huffman coding scheme is used basically. Optionally, direct natural binary coding is used to limit the worst-case number of bits as well as to correctly encode the envelope of signals which are saturated by differential Huffman coding. In order to indicate the selected mode to the spectral envelope in the decoder part, one flag bit is used. The differential Huffman coding is composed of two steps [1]. Firstly, the first index $r(0)$ is encoded by natural binary coding with 5 bits. Secondly, the differential indices and saturation flag mode are computed. Equation (2) shows the procedure to compute the differential index for the Huffman coding scheme in the lower band.

$$diff(j)=r(j)-r(j-1),\quad j=1,...,9$$
$$mode=\begin{cases}0 & if\ |diff(j)|\le 12\ for\ j=1,...,9\\ 1 & otherwise\end{cases} \tag{2}$$

where, *diff(j)* is the differential index for the j^{th} sub-band. The mode is the one-bit flag to indicate the coding scheme. If mode is 0, the Huffman coding scheme is utilized. If, mode is 1 or the total number of bits to be transmitted by differential Huffman coding is greater than 45, direct natural binary coding is used. The same procedure is applied to the higher band.

III. Proposed Algorithm based on NLMS

A. Bit Reduction for Envelope Coding based on NLMS

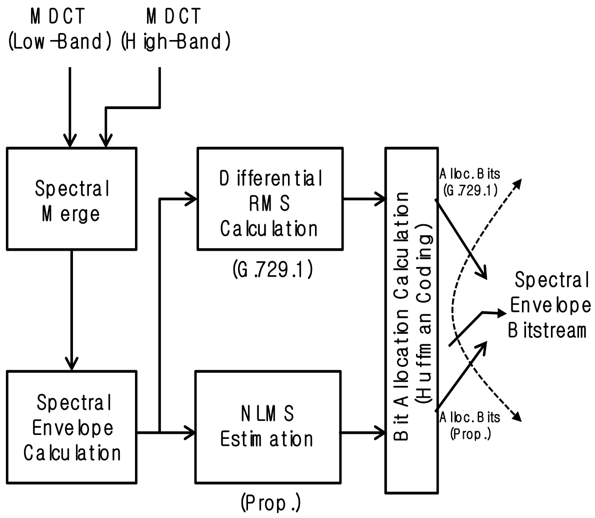

Fig. 2 Block diagram of spectral envelope coding by the proposed algorithm in the TDAC encoder

In real speech/audio data, it is natural that sub-band spectra are more correlated among adjacent consecutive frames than in a given frame. Thus, we utilize this property to reduce more bits for spectral envelope coding using NLMS. Fig. 2 represents the block diagram of spectral envelope coding by proposed algorithm based on NLMS. The procedures of spectral merge and spectral envelope calculation are same with the conventional method of G.729.1 baseline codec. In the proposed algorithm, the coding algorithms of the spectral envelope are selected between the conventional algorithm and the proposed algorithm based on NLMS through the computation of bit allocations. The one of the two algorithms that has more bit reductions is selected. Additional one bit is used to indicate which algorithm is utilized during the encoding process. NLMS algorithms are used to find the adaptive filter coefficients that minimize the error between the original sub-band energy and its estimated energy [3]. In the proposed algorithm, the estimation errors of the sub-band energy envelopes by NLMS are quantized by Huffman coding. The cost function to be minimized by NLMS is shown in (3).

$$J=E\left[\left(x(n)-\sum_{k=1}^{P}a^{(n)}(k)x(n-k)\right)^2\right]=E\left[e^2(n)\right] \tag{3}$$

where, is original envelope, and $a^{(n)}(k)$ is the k^{th} NLMS filter coefficient at the nth frame. *P* is the filter order. To simulate the filter adaptation condition in the decoder side, where exact envelope values cannot be known, the estimation errors are modified to be (4).

$$e_{\mathrm{mod}}(n)=x(n)-\sum_{k=1}^{P}a^{(n)}(k)\hat{x}(n-k) \tag{4}$$

where, $\hat{x}(n)$ is the estimated envelope in the decoder side and can be given by (5).

$$\hat{x}(n) = e_{quant}(n) + \sum_{k=1}^{P} a^{(n)}(k)\hat{x}(n-k) \quad (5)$$

where, $e_{quant}(n)$ is the quantized value of $e_{\mathrm{mod}}(n)$. The resulting filter adaptation equation by NLMS is shown in (6) [7].

$$a^{(n+1)}(k) = a^{(n)}(k) + \mu \frac{e_{quant}(n)\hat{x}(n-k)}{\|\hat{\mathbf{x}}\|^2}, \; k = 1,...,P \quad (6)$$

where, μ is filter learning rate and $\|\hat{\mathbf{x}}\| = \sum_{k=1}^{P} \hat{x}^2(n-k)$. Equations (4) to (6) are applied to quantize each sub-band envelope in the coder and the decoder side, respectively. The filter order P and the learning rate μ are pre-determined for each sub band using training data.

B. Reduced Bit Re-allocation for Sound Quality Improvement

To improve sound quality, the reduced bits by the NLMS-based envelope coding are re-allocated to reduce the quantization error of spectral envelopes. Because the error in (4) is quantized as integer units, significant sound quality degradation is inevitable. Although we can re-allocate the reduced bits to the MDCT shape, the envelope is selected to enhance because the envelope can more contributes to sound quality than the MDCT shape does [4]. To enhance the envelope more, the envelope quantization error in the high-band is vector-quantized by modified K-means (MKM) clustering. The reason why the high band is only considered is that the low-band MDCT is originated from the G.729 coding error which can be thought to have smaller quantization error than the high-band MDCT from the direct input signal.

C. Operation in Frame Error Conditions

Because the proposed envelope coding requires previous frame information, it is disadvantageous in frame error conditions. To cope with this problem, the conventional intra-frame subtraction method in G.729.1 is forcibly applied to every 5 frames. For the lost frames, $e_{quant}(n)$ in (6) is set to zero and NLMS filter coefficients are not updated.

IV. Performance Evaluation

To evaluate the performance of the proposed spectral envelope coding scheme, bit reduction, spectral distortion, and preference test score are measured. As test data, Korean spoken sentences and audio signals are used. Total numbers of speech and audio data were 20, respectively, and their lengths were around 20 seconds. The sampling rate and the resolution of the data are 16 kHz and 16 bits respectively. For frame error condition, channel model of G.191 was used [5].

A. Parameter Optimization of Adaptive Filter

The filter order P and the learning rate μ of each sub-band that minimize estimation errors are obtained by training using some speech and music data. Fig. 3 represents the example of determining the optimal learning rate and filter order of the

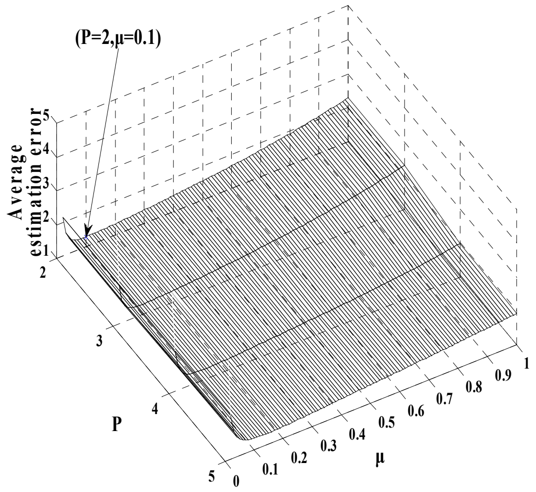

Fig. 3 Optimal learning rate for the NLMS algorithm (example of the 1st sub-band envelope)

NLMS algorithm for the 1st sub-band envelope. Table I shows optimization results of the learning rate for each sub-band. Idx is sub-band index. Throughout the sub-band envelopes, the optimal filter order is 2.

TABLE I
OPTIMAL LEARNING RATE

Idx	μ	Idx	μ	Idx	μ
1	0.10	7	0.08	13	0.06
2	0.10	8	0.06	14	0.05
3	0.08	9	0.09	15	0.07
4	0.09	10	0.06	16	0.06
5	0.06	11	0.07	17	0.07
6	0.09	12	0.05	18	0.08

B. Bit Reduction and Re-allocation

Table II shows the average numbers of reduced bits per frame for the spectral envelope coding compared with the coding by the G.729.1.

TABLE II
AVERAGE BIT REDUCTION

Test data	Bit reduction (bits)
Speech	5.72
Music	7.75

The amount of the bit allocation in G.729.1 is higher than that in the proposed algorithm. This means that the proposed algorithm based on NLMS is more efficient to reduce bits to encode spectral envelopes. As shown in the Table II, more bits are saved for music data. This reflects that the number of relatively stationary frames for audio data is generally larger than those for speech data. To improve the sound quality using the reduced bit re-allocation scheme mentioned in section III-B,

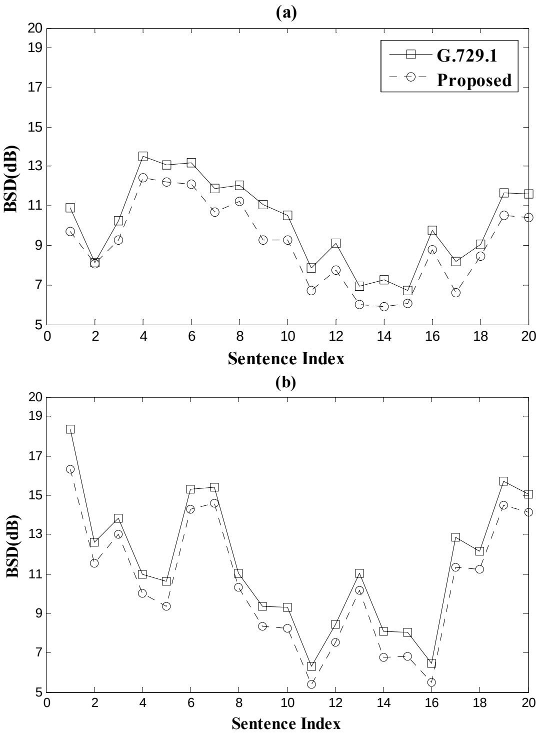

Fig. 4 Spectral distortion (BSD) result of clean DB by proposed algorithm compare with G.729.1 baseline ((a) speech DB, (b) audio DB)

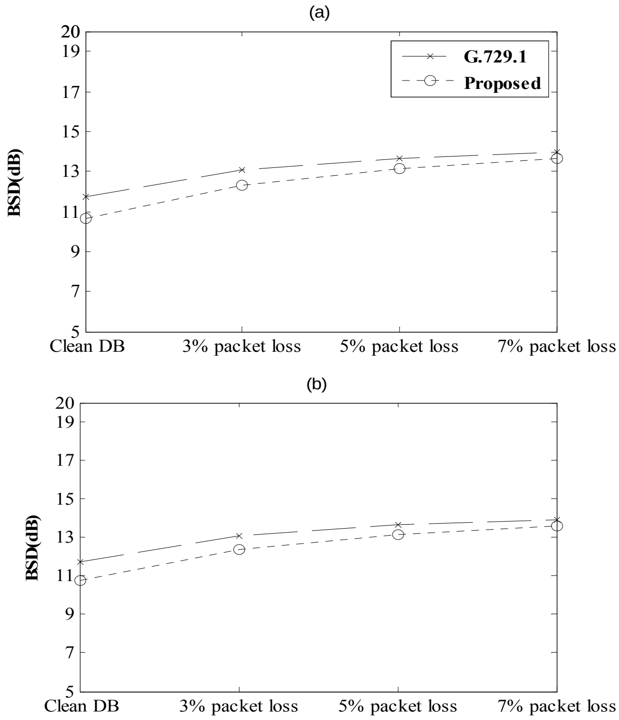

Fig. 5 Spectral distortion (BSD) result in packet loss condition by proposed algorithm compare with G.729.1 baseline ((a) speech DB, (b) audio DB)

7 bits were always used to quantize the high-band envelope by MKM clustering.

C. Sound Quality Evaluation

For the evaluation of sound quality, Bark's scale distortion (BSD) was measured between original and decoded signals [6]. In calculating the BSD, 512-FFT was performed for every 10 ms with 30 ms analysis frame. 33 energy values calculated by Bark's scale filter banks were estimated. Equation (7) represents the BSD.

$$BSD = \frac{\sum_{t=0}^{T-1}\left(BSD^{(t)}\right)}{\sum_{t=0}^{T-1}\left(\sum_{k=0}^{N-1}\left(L_x^{(t)}(k)\right)^2\right)} \tag{7}$$

where, t and k are the frame index and the filter bank index. T and N are the total numbers of frames and filter banks, respectively. $BSD^{(t)}=\sum_{k=0}^{N-1}\left(L_x^{(t)}(k)-L_y^{(t)}(k)\right)^2$. $L_x^{(t)}(k)$ and $L_y^{(t)}(k)$ are the filter bank energies of the reference and the test signal. In addition, preference tests are performed to measure subjective signal quality improvements. To measure preference scores, we tested 10 male and female listeners. Fig. 4 represents spectral distortion (BSD) result of clean DB by proposed algorithm compared with G.729.1 baseline. Table III shows preference test results. It shows that the decoded sound quality by the proposed algorithm is better than that by the conventional algorithm. Also, we can see that the performance improvement by the proposed algorithm is higher for audio data than for speech data.

TABLE III
PREFERENCE TEST SCORE (%)

Test data	G.729.1	Proposed
Speech	33.5	66.5
Music	29	71

D. Sound Quality Evaluation in Frame Error Condition

For the evaluation of sound quality in frame error conditions, the BSD is measured between original and decoded signals with frame loss of 3%, 5%, and 7%. Fig. 5 represents spectral distortion (BSD) result in packet loss condition by proposed algorithm compared with G.729.1 baseline. As shown in Fig. 5, the decoded sound qualities of the proposed algorithm are better than those of the conventional algorithm. This means that the proposed algorithm is applicable to real VoIP applications although it still has frame loss propagation.

V. CONCLUSION

This paper describes a new algorithm for encoding spectral envelope information by the NLMS adaptation error of spectral envelope in the TDAC encoder of G.729.1. The experimental results show that the proposed algorithm is efficient to reduce amount of encoding bits and to improve the quality of decoded sounds with the extra bits. Even in frame loss condition, although the proposed algorithm still has error propagation, the decoded sound quality is better than that of the conventional algorithm. The experiment results of the proposed algorithm can encourage us to apply it to real VoIP systems. As future works, studies on the method to make the spectral envelopes quantized by high resolution without additional bits will be included.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A1014222).

REFERENCES

[1] ITU-T Rec. G. 729.1, “G.729 based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729,” 2006.
[2] D. Massaloux et al, “An 8-12 kbit/s embedded CELP coder interoperable with ITU-T G.729 coder: First stage of the new G.729.1 standard,” in *Proc. ICASSP*, May 2007.
[3] Slock, D.T.M., “On the convergence behavior of the LMS and the normalized LMS algorithms”, *IEEE Transactions on Signal Processing*, Volume 41, Issue 9, pp.2811 – 2825, Sep. 1993.
[4] K. Cho, S. Jung, H. Kim, M. Hahn, “Frame error-robust MDCT bit reduction scheme using inter-frame correlation for G.729.1.,” in *Proc. ICASSP*, May 2011.
[5] ITU-T Rec. G. 191, “ITU-T Software Tool Library 2005 User's manual,” 2005.
[6] S. Wang, A. Sekey and A. Gersho, “An objective measure for predicting subjective quality of speech coders,” *IEEE Trans. Selected Areas in Comm.*, Vol. 10, No. 5, pp. 819-829, June 1992.
[7] M. Hayes, *Statistical Digital Signal Processing and Modeling*, Prentice Hall, 1996.