AN EFFICIENT TRANSCODING ALGORITHM FOR G.723.1 AND EVRC SPEECH CODERS

Kyung Tae Kim*, Sung Kyo Jung*, Young Cheol Park**, Yong Soo Choi***, and Dae Hee Youn*
* ASSP Lab. Dept. of Electric & Electronic Eng., Yonsei University,
134 Shinchon-dong, Sudaemoon-gu, Seoul 120-749, Korea
** CSPR Dept. of Electric & Electronic Eng., Yonsei University
*** LG Electronics Inc., 60-39, Kasan-dong, Kumchon-gu, Seoul 153-023, Korea

Abstract - **Interoperability is one the most important factors for a successful integration of the speech network. To operate speech networks employing different speech coders, but integrated as one, bitstreams generated by one coder should be translated seamlessly to those of the other coders. Connecting two coders in tandem may be the simplest way to accomplish this. However, coders in tandem connection often produce problems such as poor speech quality, high computational load, and additional transmission delay. In this paper, we propose an efficient transcoding algorithm that can provide interoperability to the networks employing ITU-T G.723.1 [1] and TIA IS-127 EVRC [2] speech coders. Subjective and objective quality evaluation have confirmed that the speech quality produced by the proposed transcoding algorithm was equivalent to, or better than the tandem coding, while it had shorter processing delay and less computational complexity.**

I. INTRODUCTION

Integrated speech networks often support multiple speech codecs to accommodate users with variety of qualities. In this case, interoperability between speech networks employing different speech coders becomes an important issue. For the communication between two endpoints with different speech coders, it is required to translate bitstreams generated by one coder to those of the other coder.

Two codecs can be placed in tandem to achieve this goal. Tandem is to reconstruct speech signals by decoding bitstream of one codec and to encode the speech re-constructed by another coder. However, tandem coding is associated with several problems such as: (i) Degradation of speech quality - quality degradation is inevitable because the speech signal is encoded and decoded twice using two different speech coders, (ii) High computational load - the system should implement two coders simultaneously and (iii) Long transmission delay in the communication link - tandem coding needs its own processing time as well as the look-ahead time for LPC analysis. It is clear to see that all these problems are due to the fact that the speech signal should pass through complete decoding and encoding process of the two codecs one by one. Thus, it is desirable to translate a bitstream of source coder directly into that of the target coder instead of connecting two different coders in tandem. This new procedure is called 'transcoding'.

Transcoding algorithm had been studied by others[3]. Though the previous transcoding method was suitable for coders that have the similar structure, it was not as efficient in case of two coders that have different frame/subframe size, adaptive codebook or fixed codebook. To achieve general transcoding, new algorithm needed to be researched.

This work was supported by LG Electronic Inc.

In this paper, we propose an efficient transcoding algorithm for a legitimate communication between 5.3 kbps G.723.1 used for VoIP (Voice over IP) and 8.55 kbps EVRC used for digital cellular, which are two distinctively different scheme. The proposed transcoding algorithm is composed of four parts: LSP conversion, open-loop pitch conversion, fast adaptive codebook search, and fast fixed codebook search. In LSP conversion and open-loop pitch conversion, parameters of the source coder are directly translated to the parameters of the target coder. A linear interpolation is used to translate the LSP parameters. This approach doesn't require 'lookahead', so that it can reduce the extra processing delay. In open-loop pitch conversion, open-loop pitch of the target coder is obtained from closed-loop pitch of the source coder using the pitch-smoothing method. In G.723.1 encoder, we use a new fast adaptive codebook search algorithm that uses a constraint quantization table. The proposed fast fixed codebook search method is applied to the fixed codebook search in both G.723.1 and EVRC.

Complexity checks, subjective preference tests and objective quality measures are executed to evaluate the proposed transcoding algorithm. As the result of the evaluation in comparison with the tandem coding, the proposed algorithm could reduce about 30-37% of the complexity in the encoding part, resulting significantly relieved process load and reduce in required memory, while subjective and objective quality is preserved.

II. G.723.1 & EVRC SPEECH CODERS

ITU-T G.723.1 standardized for multimedia communication speech coder, operates at two bit rates, 5.3 and 6.3 kbps. G.723.1 encodes speech or other audio signals with 30 ms frames. In addition, there is a look-ahead of 7.5 ms resulting in a total algorithmic delay of 37.5 ms. In the encoding process, for every subframe of 60 samples, a 10th order Linear Prediction Coefficients (LPC) is computed from the windowed signal. The LPC set for the last subframe is quantized using a Predictive Split Vector Quantizer (PSVQ). The unquantized LPC coefficients are used to construct the short-term perceptual weighting filter, which is used to filter the entire frame and to obtain the perceptual weighted speech signal. For every two subframes (120 samples), the open-loop pitch period is computed using the weighted speech signal. After the above processing is completed, the speech signal is then processed in adaptive codebook and fixed codebook search on a subframe basis. The adaptive codebook search is performed using the 5th order pitch predictor and the closed-loop pitch and pitch gain are computed. Finally the non-periodic component of the excitation is approximated. In fixed codebook search, two types of excitation modeling scheme are used. For the high bit rate, Multi-Pulse Maximum Likelihood Quantization (MP-MLQ), and for the low bit rate, an Algebraic Code Excited Linear Prediction (ACELP) are used, respectively.

EVRC is standardized for 8kbps speech coder in 1996 and its quality is comparable to that of 13 kbps QCELP. It has 20 ms frames and look-ahead of 10 ms resulting in a total algorithmic delay of 30 ms. EVRC speech coder is based on RCELP algorithm that employs the generalized analysis-by-synthesis scheme. Unlike conventional CELP encoders, RCELP matches time-warped version of original speech signal by a simplified pitch contour. This pitch contour is obtained via the linear interpolation of the integer pitch that is estimated in an open-loop fashion for every frame, 160 samples. While a fractional pitch is transmitted in conventional CELP, the integer pitch is transmitted over each frame of speech in RCELP. Using a rate determination algorithm, EVRC encoder selects one out of three encoding rates: Rate 1 (8.55 kbps), Rate 1/2 (4 kbps), and Rate 1/8 (800 bps). At Rate 1/2 and Rate 1, the encoder uses the RCELP algorithm to match a time-warped version of the original speech. At Rate 1/8, the encoder just characterizes its energy contour. The FCB in EVRC has an algebraic codebook structure designed using an Interleaved Single-Pulse Permutation (ISPP) scheme.

III. The Proposed Transcoding Algorithm For G.723.1 & EVRC

A. Transcoding from G.723.1 to EVRC

Proposed transcoding algorithm from G.723.1 encoder to EVRC decoder consists of LSP conversion using linear interpolation, open-loop pitch conversion using pitch smoothing and fast fixed codebook search algorithm. By considering the frame length of two speech coders, two frames of G.723.1(30 ms) are translated to three frames of EVRC corresponding to 20 ms each. A block diagram of the developed transcoding algorithm from G.723.1 encoder to EVRC decoder is shown in Fig. 1.

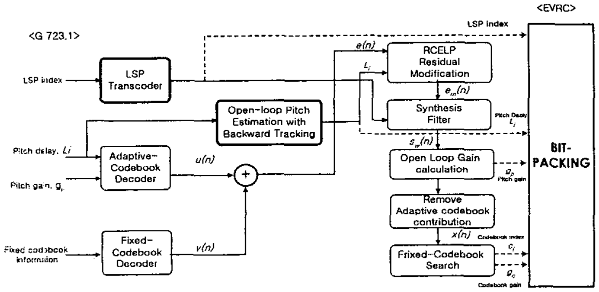

Fig. 1: Block diagram of the transcoding from G.723.1 to EVRC

LSP conversion using linear interpolation

Linear interpolation was used to translate two set of LSP information of G.723.1 into three sets of LSP parameters of EVRC. The LSP conversion procedure is written by

$$
\begin{aligned}
P_{C1}(j) &= 0.5417P_{A10}(j) + 0.4583P_{A11}(j) \\
P_{C2}(j) &= 0.8750P_{A11}(j) + 0.1250P_{A12}(j) \quad ,1 \le j \le 10 \\
P_{C3}(j) &= 0.2083P_{A11}(j) + 0.7917P_{A12}(j)
\end{aligned} \tag{1}
$$

where P_{Ai} and P_{Ci} are LSP of G.723.1 and EVRC respectively, and i is frame index. The LSP conversion does not require the calculation of auto-correlation coefficient, Durbin Recursion for LPC and the conversion LPC to LSP resulting relatively lower computational complexity than of tandem coding.

In EVRC, the lookahead is need for modification of the residual as well as LPC analysis, but that process does not need all lookaheads. When the residual is modified, the pulses in the original residual are shifted by less than 30 samples. Consequently the proposed method is to use only 30 samples of the lookahead. We can get the same result as original EVRC, though using only 30 samples of the lookahead. Therefore, the algorithmic delay of transcoding results 6.25 ms decrease from that of tandem coding.

Open- loop pitch estimation

After LSP conversion process, the open-loop pitch of EVRC is computed using the closed-loop pitch of G.723.1 in the weighted speech domain. The open loop pitch estimation is performed around the closed-loop pitch of corresponding G.723.1 subframe.

In the proposed transcoding algorithm, the open-loop pitch estimation process is simplified using the pitch smoothing scheme. The closed-loop pitch of G.723.1 is compared with the one from the previous EVRC subframe. If the distance of two pitch values is less than 10 samples, considering the similarity of two pitch values, the closed-loop pitch of G.723.1 is determined as the open-loop pitch of EVRC. Otherwise, the pitch smoothing method is applied. In the pitch smoothing method, local maximum delay which maximizes Eq. (2) in the range of ± 3 sample boundary around the closed loop pitch of G.723.1 and EVRC:

$$
R(k_i) = \sum_{n=0}^{79} s_w(n) \cdot s_w(n-k_i) , \begin{cases} p_A - 3 \le k_1 \le p_A + 3 \\ p_B - 3 \le k_2 \le p_B + 3 \end{cases}, \tag{2}
$$

where $s_w(n)$ is the weighted speech signal, p_A and p_B are closed-loop pitch of G.723.1 and EVRC, respectively. And k_1 and k_2 are open loop pitch candidates of corresponding range. After determining the local maximum delay which is maximizing $R(k_i)$ in each range, $R(k_i)$ is normalized by the energy at the local maximum delay:

$$
R'(D_i) = \frac{R(D_i)}{\sqrt{\sum_{k=0}^{159-D_i} \varepsilon^2(k) \sum_{k=0}^{159-D_i} \varepsilon^2(k+D_i)}} , \quad i = 1, 2 \tag{3}
$$

where D_i is local maximum delay at each range, D_1 and D_2 are the local maximum delays of G.723.1 and EVRC, respectively.

Then, the normalized local maximum values are compared with each other, with more weighting on EVRC. The smoothed open-loop pitch is determined as

$$
\begin{aligned}
& if\ [\{(D_2 > D_1 + 6)\ \&\ (\beta_1 > 0.6\beta_2)\} \\
& \quad \|\ \{(D_2 < D_1 - 6)\ \&\ (\beta_1 > 1.2\beta_2)\}] \\
& \qquad \text{delay} = D_1 \\
& else \\
& \qquad \text{delay} = D_2
\end{aligned}
$$

where D_1 and β_1 are the pitch and the gain at the previous subframe of EVRC and D_2 and β_2 are the pitch and the gain at the G.723.1.

Fast fixed codebook search for EVRC

The fixed codebook is based on an algebraic codebook structure, which has advantages in terms of storage, search complexity, and robustness. In Rate 1 fixed codebook, every codebook vector of length 55 contains at most 8 non-zero pulses. The 55 positions in a subframe are divided into 5 tracks(T0, T1, T2, T3 and T4). The 8 pulses are grouped into 4 pairs of pulses. The pulse positions shall be determined sequentially one pair at a time. In the first iteration, the single-pulse tracks are T3 and T4, the search process shall be repeated for other 3 iterations, by assigning the single-pulse tracks to T4-T0, T0-T1, and T1-T2 respectively. As a result, there is four codewords corresponding to four sets of the single-pulse track.

In the proposed algorithm, the searching codeword is selected in advance and only one codeword is searched. Criterion for selecting the searching codeword is the energy of the backward filtered signals in the single-pulse track. The codeword that has the minimum energy is selected and only the selected codeword is searched. This method reduces the combination of pulse location by 1/4 and gives equivalent quality to the original one.

B. Transcoding from EVRC to G.723.1

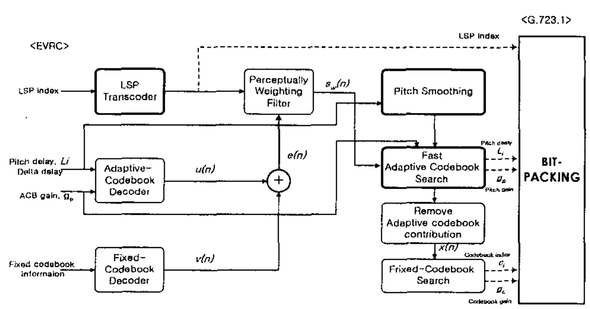

Fig. 2 Block diagram of the transcoding from EVRC to G.723.1

For the case of speech transmission from EVRC encoder to G.723.1 decoder, the LSP conversion using linear interpolation, open-loop pitch conversion using pitch smoothing, fast adaptive codebook search and fast fixed codebook search [3] schemes are used. Parameters corresponding to three frames of EVRC are converted to parameters corresponding to two frames of G.723.1. Transcoding structure in this case is shown in Fig. 2.

LSP conversion and open-loop pitch conversion is similar to the case of the transcoding from G.723.1 to EVRC. Fast adaptive codebook search and fast fixed codebook search is described in 3.2.1 and 3.2.2, respectively.

Fast adaptive codebook search

In G.723.1, the adaptive codebook search uses a 5th order pitch predictor. This process is computationally demanding because the pitch delay and pitch gain are searched simultaneously. Previously, we proposed a fast adaptive codebook search algorithm [5]. In this algorithm, pitch delay and pitch gain are computed sequentially. At first, the pitch delay is computed using a 1st order pitch predictor, and later, the pitch gains of the 5th order pitch predictor are computed. This algorithm enables the system to save a significant computational power [5].

Vector quantization of the pitch gain of 5.3 kbps G.723.1 coder uses 170 entries codebook. This process is another major computational burden for the system. In the developed transcoding algorithm, the search range of gain codebook is limited by the pitch gain of EVRC. Previously, the overall gain of each 170 entries is calculated. 170 entries are classified into the 8 different sets of range by the overall gain and the only one set of range selected by the pitch gain of EVRC is searched. Experimental test results validated that the proposed method does not reduce the quality perceptually.

Fast fixed codebook search for G.723.1

In the fixed codebook search of G.723.1, 4 pulses are searched based on ACELP structure for every subframe. Each subframe is divided by 4 tracks, and the pulse and sign of each pulse are determined using nested-loop search. As a result, the pulse locations are searched with the combination of 8^4 in the theoretically worst case. Practically, limiting the number of entering the loop for the last pulse search reduces the complexity.

In this paper, the depth-first tree search is used for fixed codebook search of G.723.1. The combination of pulse location, consequently, can be reduced up to 1/16 in the optimal case.

IV. EVALUATIONS

A. Objective quality evaluation

LPC-CD (LPC Cepstral-Distance) and PSQM [6] are used for objective evaluation measures. 8 seconds-long sentences were recorded with two male and female speakers, and the speech was sampled at 8kHz. The results for the tandem coding and the transcoding are shown in Table 1. As shown in Table 1, LPC-CD and PSQM of the trancoding indicate lower values than the tandem coding.

B. Subjective quality evaluation

An informal A-B preference test was conducted for a subjective evaluation involving 20 listeners. In this test, the subjects had to make a forced choice between pairs of samples presented over headphone set. The test material included 4 clean speech sentences composed of two male and two female speakers each. Table 2 shows the result of blind A-B preference test. As shown in Table 2, the ratio of the preferring the tandem and transcoding is similar. Results imply that the listeners could not distinguish the quality of the tandem coding from that of the transcoding. Thus, it can be inferred that the proposed transcoding algorithm produces the speech with quality equivalent to that of tandem coding.

C. Complexity check

To compare the complexity of the tandem coding with transcoding, we have implemented two algorithms into TMS320C6201 DSP and checked the cycles for encoding. As shown in Table 3, the processing cycles of the each module employing the transcoding algorithm is noticeably decreased. It is observed that the transcoding algorithm presents the total complexity about 30-37% lower than the tandem coding.

TABLE 1
OBJECTIVE EVALUATIONS

	LPC-CD(dB)		PSQM	
	Female	Male	Female	Male
Tandem(AtoC)	5.51	4.46	3.21	2.65
Transcoding(AtoC)	5.14	3.56	2.97	2.35
Tandem(CtoA)	4.94	3.95	3.00	2.41
Transcoding(CtoA)	4.76	3.17	3.00	2.35

(AtoC(from G.723.1 to EVRC), CtoA(from EVRC to G.723.1))

TABLE 2
SUBJECTIVE PREFERENCE

	G.723.1→EVRC		EVRC→G.723.1	
	Female	Male	Female	Male
Tandem	40 %	40 %	35 %	40 %
Transcoding	40 %	55 %	40 %	55 %
No Preference	20 %	5 %	25 %	5 %

TABLE 3
COMPLEXITY

	G.723.1 → EVRC		EVRC → G.723.1	
	Tandem	Transcoding	Tandem	Transcoding
LPC/LSP	930998	249303	306794	215881
Open-loop pitch estimation	263564	132238	213682	132302
Adaptive codebook search	620584	506014	826792	462816
Fixed codebook search	2104629	971666	624888	170008
Others	2954342	2954342	920985	920985
Total	6874118	4813563	2893142	1901992

V. CONCLUSION

In this paper, we proposed an efficient transcoding algorithm for 5.3 kbps G.723.1 and 8.55 kbps EVRC. This transcoding algorithm, without a noticeable loss at the subjective quality degradation, reduces both the complexity and the delay time, all which tandem coding fails to provide. The proposed transcoding algorithm provides equivalent speech quality to the tandem coding with the shorter delay and less computational complexity.

The proposed algorithm is also applied to transcoding between CELP-type speech coders such as G.729 and AMR (Adaptive Multi-Rate) speech coder.

VI. REFERENCES

[1] ITU-T Rec. G.723.1 "Dual-rate Speech Coder For Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s," 1996.

[2] TIA/EIA/IS-127 Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems, 1996.

[3] H. G. Kang, H. K. Kim, and R. V. Cox, "Improving Transcoding Capability of Speech Coders in Clean and Frame Erased Channel Environments," proc. IEEE Workshop on Speech coding, pp. 78-80, 2000

[4] Sung Wan Yoon, Sung Kyo Jung, Young Cheol Park, Yong Soo Choi, and Dae Hee Youn, "A Study on Transcoding Algorithm for G.723.1 and G. 729A Speech Coders," EUROSPEECH 2001, to be appeared

[5] Sung-Kyo Jung, Young-Cheol Park, Sung-Wan Youn, Kyoung-Tae Kim, and Dae-Hee Youn, "Efficient Implementation of ITU-T G.723.1 Speech Coder for Multichannel Voice Transmission and Storage," EUROSPEECH 2001, to be appeared.

[6] ITU-T Rec. P. 861 "Objective Quality Measurement Of Telephoneband(300–3400Hz) Speech Codecs," 1996.