AN EFFICIENT TRANSCODING ALGORITHM FOR G.723.1 AND G.729A SPEECH CODERS

Sung Wan Yoon, Sung Kyo Jung*, Young Cheol Park**, and Dae Hee Youn**

*ASSP Lab. Dept. of Electric & Electronic Eng., Yonsei University,
134 Shinchon-dong, Sudaemoon-gu, Seoul 120-749, Korea
** CSPR, Yonsei University,134 Shinchon-dong, Sudaemoon-gu, Seoul 120-749, Korea
yocello@assp.yonsei.ac.kr

Abstract

To set a valid communication channel between two endpoints employing different speech coders, decoder and encoder of each endpoint need to be placed in tandem. However, tandem coding is often associated with problems such as poor speech quality, high computational load, and additional transmission delay. In this paper, we propose an efficient transcoding algorithm for a legitimate communication between 5.3 kbps G.723.1 and 8 kbps G.729A coders. 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 each part of the transcoding algorithm, parameters of the target coder are obtained directly from the parameters of the source coder. The efficient transcoding algorithm is supported via the computational reduction of about 25-35% in the encoding part. Subjective preference tests as well as objective quality evaluation confirmed that the proposed transcoding algorithm can produce equivalent speech quality to the tandem coding with the shorter processing delay and less computational complexity.

1. Introduction

Variety of speech coding standards have been established over last decades. Among them, G.723.1 and G.729 cover a wide range of applications with low bit rate requirements. Each standard may have different applications due to their operational characteristics. However, for some applications such as digital cellular and Voice over IP (VoIP), it is required to support both, or even more, standards to manifest interoperability. In such cases, the system is often encountered with having to set speech communication between two endpoints employing different type of coders. A simple solution to this problem is to place decoder/encoder of one endpoint and encoder/ decoder of the other endpoint in tandem, as shown in Fig 1(a)

Tandem coding is associated with several problems such as

- Degradation of speech quality - quality degradation is inevitable because the speech signal is encoded and decoded twice using two different speech coders.
- High computational load – the system should implement two coders simultaneously.
- Long transmission delay in the communication link - tandem coding needs the processing plus the look-ahead data samples for LPC analysis.

It is clear to see that all these problems are due to fact that the speech signal should pass through complete process of the two speech coders in tandem. Thus, it is desirable to translate a bitstream of source coder into that of the target coder, called *transcoding*. By doing so, it is possible to prevent performance degradation because source parameters are directly translated into target ones instead of being re-estimated from the decoded PCM data. Also, the processing delay and the computational complexity can be reduced. Fig. 1(b) shows a block diagram of the transcoding. In Fig. 1(b), the speech signal is decoded only one time.

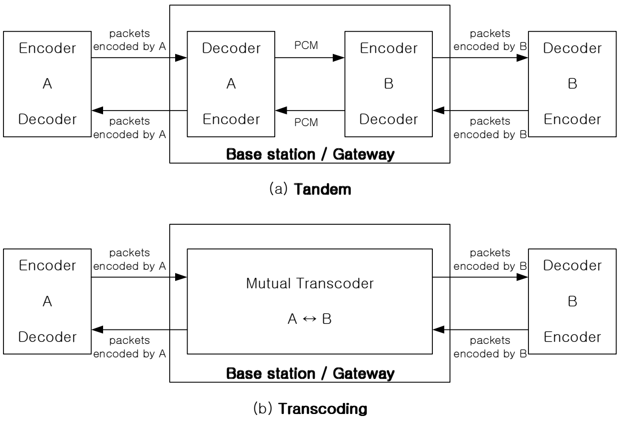

Fig. 1 (a) Tandem (b) Transcoding

In this paper, we propose an efficient transcoding algorithm between two endpoints working with 5.3 kbps G.723.1 [1] and 8kbps G.729A [3] speech coders, respectively. The proposed transcoding algorithm is composed of four parts: LSP conversion, open loop pitch conversion, fast adaptive codebook search, and fast fixed codebook search. By considering the frame length of two-speech coders, parameters corresponding to one frame of G.723.1 are converted to three sets of equivalent G.729A parameters. In addition to a complexity measure, the performance was evaluated via subjective preference tests as well as objective quality evaluations including LPC-CD, PSQM with various speech sets.

2. ITU-T G.723.1 & G.729A 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 msec frames. In addition, there is a look-ahead of 7.5 msec resulting in a total algorithmic delay of 37.5 msec.

In the encoding process, for every subframe of 60 samples, 10th order Linear Prediction Coefficients (LPC) are computed from the windowed signal. The LPC set for the last subframe

This work was supported by LG Electronics Inc.

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 subframe (120 samples), the open-loop pitch period is computed using the weighted speech signal. After the above processing, the speech signal is 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) is used, respectively.

ITU-T G.729 [2] based on CS-ACELP (Conjugated-Structure ACELP) operates at 8kbps. G.729 encodes speech or other audio signals in 10 msec frames and there is additional look-ahead of 5 msec, resulting in a total algorithmic delay of 15 msec.

In the encoder, for every frame of 80 samples, a 10th order LPC filter is computed using the Levinson-Durbin recursion. The LPC filter for the 2nd subframe is quantized using a Multi-Stage Vector Quantization (MSVQ). 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. After the computing the weighted speech signal, the open-loop pitch period is computed. To avoid choosing the pitch multiples, the open-loop pitch estimation procedure divides the delay range onto three sections and favoring the smaller values. And then, adaptive codebook and fixed codebook search is performed on a subframe basis. The adaptive codebook search is performed using the 1st order pitch predictor. The fractional pitch delay is searched with 1/3 resolution. In the fixed codebook search, non-periodic component of excitation is modeled by ACELP using 4 pulses. For the efficiency of quantization process of pitch gain and fixed codebook gain, the two codebooks of conjugate structure are used.

G.729 Annex A coder is a complexity-reduced version of G.729 [3]. The complexity of G.729A is about 50% of G.729. The bit allocation is the same as that of original G.729. The major algorithmic changes to the full version of G.729 are perceptual weighting filter, open-loop pitch estimation, adaptive codebook search, fixed codebook search and the post filter parts.

3. The Proposed Transcoding Algorithm

3.1. From G.723.1 to G.729A

The proposed transcoding algorithm has asymmetric structure for Tx (Transmission) and Rx (Receive) paths. For the transmission of speech data from G.723.1 encoder to G.729A decoder, transcoding process involves LSP conversion using linear interpolation and open-loop pitch conversion using pitch smoothing. By considering the frame length of two speech coders, one frame of G.723.1 corresponding to 30 msec is translated to three frames of G.729A corresponding to 10 msec each. A block diagram of the developed transcoding algorithm from G.723.1 encoder to G.729A decoder is shown in Fig. 2. As shown in Fig. 2, the left dotted box is a decoder module of G.723.1 and the right one is an encoder module of G.729A.

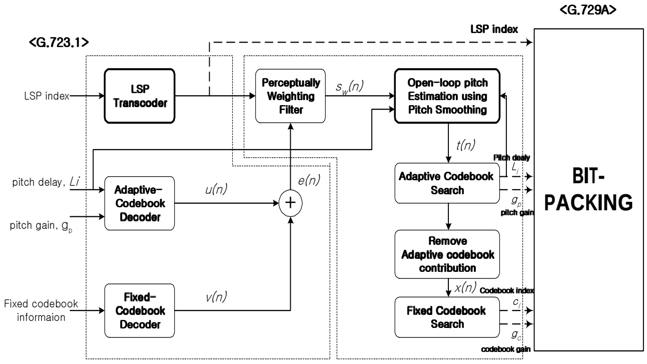

Fig. 2 Block diagram of the transcoding from G.723.1 to G.729A

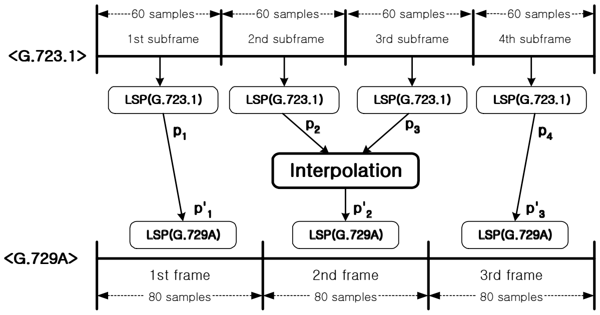

Fig. 3 LSP conversion using linear interpolation

3.1.1. LSP conversion using linear interpolation

Linear interpolation was used to translate the LSP information of each subframe of G.723.1 into three sets of LSP parameters of G.729A. The LSP conversion procedure shown in Fig. 3 is written by

$$p_i^{'}(j) = \begin{cases} p_i(j), & i = 1 \\ \dfrac{1}{2}\left(p_i(j) + p_{i+1}(j)\right), & i = 2 \\ p_{i+1}(j), & i = 3 \end{cases} \quad ,1 \le j \le 10, \qquad (1)$$

where p_i and p_i' are LSP of G.723.1 and G.729A respectively, and i is frame index.

LSP conversion process can be also applied to the case of speech transmission from G.729A encoder to G.723.1 decoder. Fig. 4 shows the LPC spectrum of the voiced region of speech signal. The LPC spectrum of G.729A, which is the target Rx decoder, is used as a reference. As shown in Fig. 4, the LPC spectrum obtained after transcoding matches closely to the reference spectrum in low frequency and formant region having a much effect on the speech quality. LPC spectrum with tandem coding, however, indicates more spectral distortion than transcoding. Since the proposed transcoding is not involved with the LPC analysis and LSP conversion, it can reduce the overall complexity. In tandem coding, additional 5 ms look-ahead is needed for LPC analysis. But this look-ahead delay is not necessary in transcoding because the LPC analysis is not required after all.

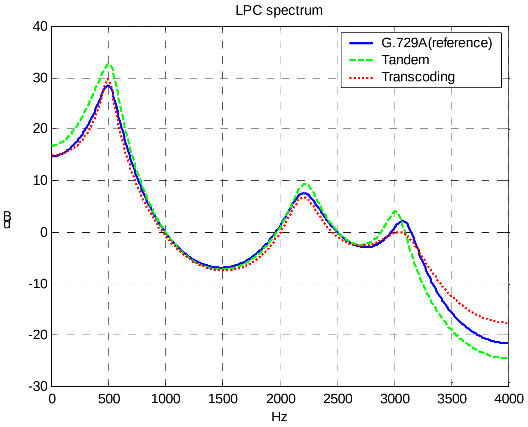

Fig. 4 LPC spectrum : reference(solid), tandem(dash), transcoding(dot)

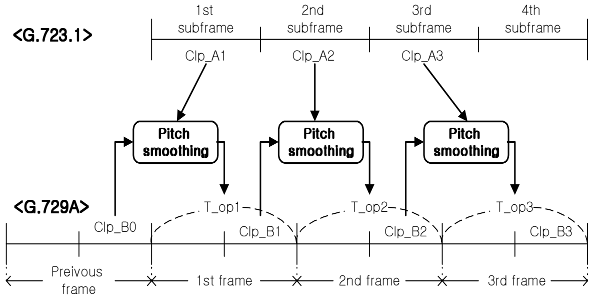

Fig. 5 Open-loop pitch estimation using pitch smoothing

3.1.2. *Open-loop pitch estimation using pitch smoothing*

After LSP conversion process, the open-loop pitch of G.729A 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 as shown in Fig. 5. The closed-loop pitch of G.723.1 is compared with the one from the 2nd subframe of the previous G.729A frame. 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 G.729A. 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 G.729A:

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

where $s_w(n)$ is the weighted speech signal, p_A and p_B are closed-loop pitch of G.723.1 and G.729A, 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:

$$R'(t_i)=\frac{R(t_i)}{\sqrt{\sum_n s_w^2(n-t_i)}}\,,\quad i=1{,}2\quad, \qquad (3)$$

where t_i is local maximum delay at each range, t_1 and t_2 are the local maximum delays of G.723.1 and G.729A, respectively.

Later, the normalized local maximum values are compared each other with more weighting on G.729A. That is, if the local maximum value of G.729A is lager than 3/4 times of that of G.723.1, the open-loop pitch of G.729A is determined as the local maximum delay of G.729A. Otherwise, the local maximum delay of G.723.1 is selected. The smoothed open-loop pitch, T_{op}, is determined as

```
$T_{op}=t_1$
$R'(T_{op})=R'(T_1)$
$if\ R'(t_2)\ge 0.75\cdot R'(T_{op})$
        $R'(T_{op})=R'(T_2)$
        $T_{op}=t_2$
$end.$
```

Because the autocorrelation of the weighted speech is not computed and the maximum value search process is not involved, it is possible to estimate the open-loop pitch with much less computational load. Also, the pitch smoothing scheme can reduce the quantization noise due to the inaccuracy of pitch. Thus, it is expected that the speech quality could be improved or, at least, comparable to that of tandem coding.

3.2. From G.729A to G.723.1

For the case of speech transmission from G.729A 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 G.729A are converted to parameters corresponding to one frame of G.723.1. Transcoding structure in this case is similar to the structure shown in Fig. 2. However, there are couple of modules added to the structure, which are a fast adaptive codebook search module and a fast fixed codebook search module.

3.2.1. *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 G.729A. Thus by the distribution of pitch gain of G.729A, the search range of gain codebook is limited to the pre-selected 85 entries.

3.2.2. Fast fixed codebook search

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 84, in the theoretically worst case, using analysis-by-synthesis. 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, is reduced to $2\times\{(8\times8)+(8\times8)\}$.

4. Evaluations

4.1. Objective quality evaluation

LPC-CD (LPC Cepstral-Distance) and PSQM [6] are used for objective evaluation measures. 8 sec sentences were recorded with two male and female speakers under quiet environment, 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. The results can be judged as the transcoding can provide better subjective quality to the listener.

4.2. Subjective quality evaluation

An informal A-B preference test was conducted for a subjective evaluation involving 30 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.

4.3. Complexity

To compare the complexity of the proposed algorithm, both tandem and transcoding algorithm were implemented on TI TMS320C6201 DSP chip. Because we focused on just comparing the complexity, the optimization process in the implementation was omitted. So the figures in Table 3 are not the optimal results on the view of the DSP implementation. But it makes no difference on the objective complexity performance of tandem and transcoding algorithm because the complexity in encoding part differs only in which transcoding algorithm is applied.

As shown in Table 3, the processing time of the each module employing the transcoding algorithm is noticeably decreased. Also, the total encoding time of the transcoding is decreased to the level of 63-74% of the tandem coding. Thus, it can be said that the transcoding algorithm can synthesize a speech of the quality equivalent to the tandem coding with complexity about 26-37% lower than the tandem coding.

Table 1 Objective quality evaluation

		LPC-CD(dB)		PSQM	
		Male	Female	Male	Female
G.723.1 ➔ G.729A	Tandem	3.90	3.98	2.44	2.45
	Transcoding	3.54	3.66	2.17	2.22
G.729A ➔ G.723.1	Tandem	3.65	4.17	2.43	2.47
	Transcoding	3.25	3.86	2.27	2.46

Table 2 Subjective preference

Preference	G.723.1 ➔ G.729A		G.729A ➔ G.723.1	
	Female	Male	Female	Male
Tandem	30 %	20 %	26.7 %	30 %
Transcoding	36.7 %	33.3 %	13.3 %	40 %
No Preference	33.3 %	46.7 %	60 %	30 %

Table 3 Complexity check using TMS320C6201

MIPS	G.723.1 ➔ G.729A		G.729A ➔ G.723.1	
	Tandem	Transcoding	Tandem	Transcoding
LPC	6.41	2.36	6.93	5.55
Olp	0.94	0.21	1.54	1.19
ACB	2.45	2.45	10.14	6.34
FCB	4.30	4.30	10.50	2.16
Others	4.04	4.04	8.05	8.05
Total	18.15	13.37	37.17	23.28

5. Conclusion

In this paper, we proposed an efficient transcoding algorithm that could translate 5.3 kbps G.723.1 bitstream into 8 kbps G.729A bitstream. This transcoding algorithm is appropriate to prevent the problems arising when we use a simple tandem coding technique, such as quality degradation, high complexity, and increase of the delay time. The proposed transcoding algorithm is composed of four parts: LSP conversion, open-loop pitch conversion, fast adaptive codebook search, and fast fixed codebook search. Results of subjective and objective evaluation showed that the proposed transcoding algorithm can produce equivalent speech quality to the tandem coding with the shorter delay and less computational complexity.

6. 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] ITU-T Rec. G.729 “Coding of Speech at 8 kbit/s CS-ACELP Speech Coder,” 1996.

[3] ITU-T Rec. G.729 Annex A “Reduced Complexity 8 kbit/s CS-ACELP Speech Codec,” 1996.

[4] S.W. Youn S.K. Jung, Y.C. Park, and D.H. Youn, “Transcoding Algorithm from 8 kbps G.729A to 5.3 kbpsG.723.1”, *Proc. KSPC*, pp. 823-826, Sep. 2000.

[5] S.K. Jung, Y.C. Park, S.W. Youn, I.H. Cha, and D.H. Youn, “A Proposal of Fast Algorithms of ITU-T G.723.1 for Efficient Multi channel Implementation”, *Proc. KSCSP*, pp67-70, 2000.

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