A Novel Transcoding Algorithm for SMV and G.723.1 Speech Coders via Direct Parameter Transformation

Seongho Seo, Dalwon Jang, Sunil Lee, and Chang D. Yoo

Department of Electrical Engineering and Computer Science,
Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

dansoc@mail.kaist.ac.kr

Abstract

In this paper, a novel transcoding algorithm for the Selectable Mode Vocoder (SMV) and the G.723.1 speech coder is proposed. In contrast to the conventional tandem transcoding algorithm, the proposed algorithm converts the parameters of one coder to the other without going through the decoding and encoding process. The proposed algorithm is composed of four parts: the parameter decoding, Line Spectral Pair (LSP) conversion, pitch period conversion and rate selection. The evaluation results show that the proposed algorithm achieves equivalent speech quality to that of tandem transcoding with reduced computational complexity and delay.

1. Introduction

Today, there exists a variety of wire and wireless communication networks in which different speech coding standards are adopted. With the availability of Internet growing, the Voice over Internet Protocol (VoIP) service across various networks and the interoperability of different communication networks are becoming a concern. Since each network uses a different speech coder, one is not compatible with others. To address this problem, a transcoding algorithm, which is to convert the encoded bit stream of one speech coder into that of the other, is required. In this paper, a transcoding algorithm for Selectable Mode Vocoder (SMV)[2] and G.723.1[1] speech coder is proposed. The SMV is a new speech coding standard for CDMA2000 system. On the other hand, G.723.1 speech coder is currently widely used for VoIP service. Generally, a simple solution to the compatibility problem is to decode the bit stream of one coder and then encode the generated speech using the other speech coder. This method is called tandem transcoding and is associated with a number of problems such as the degradation of speech quality and the increases in computational complexity and delay. These problems can be alleviated by the proposed transcoding algorithm which converts the parameters of one coder to the other without going through the decoding and encoding processes. This paper proposes a novel transcoding algorithm for SMV and G.723.1 speech coder via direct parameter transformation. In [4]-[6], transcoding algorithms for other combination of speech coders are evaluated.

The rest of this paper is organized as follows. Section 2 introduces the SMV and G.723.1 speech coders. In Section 3 and Section 4, the proposed transcoding algorithm is described in detail. Section 5 evaluates the performance of the proposed method. Finally, Section 6 concludes the paper.

2. SMV and G.723.1 Speech Coders

2.1. G.723.1 Speech Coder

G.723.1 speech coder operates at two bit rates, 5.3 and 6.3 kbit/s. The coder takes 16 bit linear PCM sampled at 8kHz as input. The length of each speech frame is 30 ms which corresponds to 240 samples. Input speech of each frame is first high pass filtered and then divided into 4 subframes of 60 samples each. For every subframe, a 10th order linear prediction coefficients (LPC) are computed. The linear predictive analysis requires a look-ahead of 7.5 ms long. The LPC set for the last subframe is quantized using the Predictive Split Vector Quantizer (PSVQ) and transmitted. The unquantized LPC sets are used to obtain the perceptually weighted speech signal. For every two subframes, the open-loop pitch analysis is performed in the domain of the weighted speech signal. Then, the adaptive codebook (ACB) and fixed codebook (FCB) are searched on a subframe basis. In the ACB search, the closed-loop pitch period and ACB gain are computed using the open-loop pitch period and the 5th order pitch predictor. Finally, the non-periodic component of the excitation is approximated using the fixed codevector. All processes except FCB search are the same for two operating rates. For the high rate, the Multi-pulse Maximum Likelihood Quantization (MP-MLQ), and for the low rate, the Algebraic Code-Excited Linear Prediction (ACELP) are used in the FCB search, respectively.

The G.723.1 speech coder employs the silence compression scheme to reduce average bit rate by an external option. This scheme includes a voice activity detection (VAD) algorithm and a comfort noise generation (CNG) algorithm. G.723.1 detects a silence interval of speech using VAD and encodes those speech frames with a very low bit rate, which is lower than 5.3kbps, using CNG algorithm.

2.2. SMV Speech Coder

The SMV operate at four different rates: 8.55, 4.0, 2.0, and 0.8 kbit/s. These rates are called Rate 1, 1/2, 1/4, and 1/8 respectively. In addition, SMV has 4 network-controlled operating modes. The different modes allow a tradeoff between average data rate (ADR) and speech quality. A 20ms frame of speech signal sampled at 8kHz, which corresponds to 160 samples, is processed at one of four rates. A rate for each frame is determined by a rate-determination algorithm (RDA). The rate selection is based on the frame classification of each frame (voiced speech, unvoiced speech, background noise, stationarity, etc.) and controlled by the SMV mode.

The SMV speech coder encodes the input speech as follows. First, the input speech is pre-processed. The pre-processing includes the silence enhancement, high-pass filtering, noise suppression, and adaptive tilt compensation. Second, frame processing that includes LPC analysis, open-loop pitch, search, signal modification, and classification are performed . The LPC analysis is performed three times for each frame and each analysis uses a different weighting window. Only one LPC set by the analysis with the window centered on the last quarter of the frame is converted to LSP, quantized and transmitted. Other sets are used for other processing such as
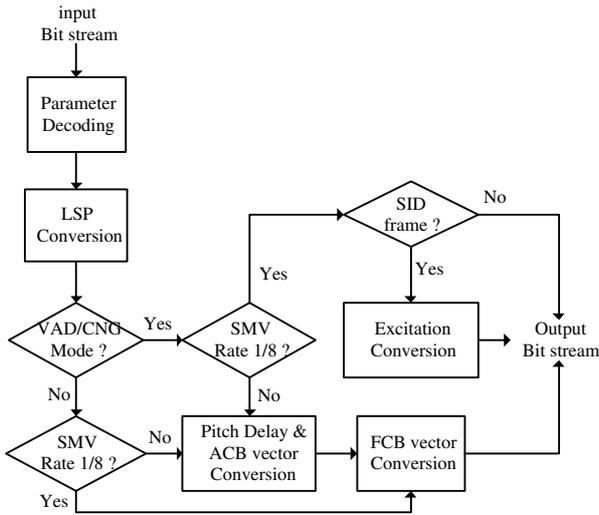
Figure 1: *Block diagram of transcoding algorithm from SMV to G.723.1*

Figure 2: *LSP conversion via linear interpolation (SMV → G.723.1)*

generation of the weighted speech and the frame classification. For LPC analysis, a lookahead of 10ms is needed. For every half frame, the open-loop pitch period is computed using the weighted speech signal. In the signal modification procedure, the weighted speech is warped to match the pitch contour calculated from open-loop pitch period. In the classification, the current frame is classified as either silence, noise-like unvoiced, unvoiced, onset, non-stationary voiced or stationary voiced and one of 4 possible rates is selected for that frame, according to the mode and the frame class. In addition, every frame selected as Rate 1 or Rate 1/2 is declared as Type 0 or Type 1. The Type 1 frames are frames of stationary voiced speech, while frames of Type 0 are all other types of speech. Type 1 frames assign more bits to FCB and less bits to ACB than Type 0 frames. Third, depending on the rate, SMV calculates the excitation of the synthesis filter differently. For Rate 1/4, the excitation signal is generated by a random number generator, multiplied by a gain factor at each subframes of 2 ms, and then filtered by a selected frequency-shaping filter. For Rate 1/8, the excitation signal is also generated randomly and then multiplied by a single gain for the frame. For Rate 1 and Rate 1/2, the calculation of the excitation is based on the extend CELP (eX-CELP)[3].

3. Transcoding from SMV to G.723.1

3.1. General Description

The SMV and G.723.1 speech coder have different frame length. Three frames of SMV corresponding to 60ms is translated to two frames of G.723.1. Fig. 1 shows a block diagram of the proposed transcoding algorithm going from the SMV to G.723.1. In the parameter decoding part, the parameters to be transcoded are decoded from the input bit-stream of the SMV. The parameters are LSP, pitch period for ACB and encoding rate of the SMV. After the parameters are decoded, the transcoder converts the LSP of the SMV into that of the G.723.1 using linear interpolation.

The Rate 1/8 of the SMV is available for frames classified either as background noise or silence. Since the adaptive codevector represents the periodic characteristic of the speech, the ACB search is not needed for the frame of G.723.1 corresponding to the Rate 1/8 frame of SMV. In this rate, the proposed algorithm performs only FCB search to obtain the excitation
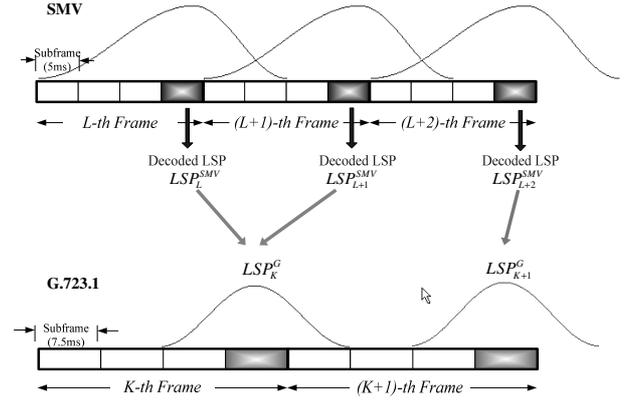
signal for the LPC synthesis filter. In other rates, fixed and adaptive codevector are converted one after another. Since the G.723.1 performs the silence compression by the external option or VAD/CNG option, the proposed algorithm also works in the special procedure including CNG by VAD/CNG option. If the external option indicates the use of the silence compression scheme and the SMV operates in Rate 1/8, then the proposed algorithm performs CNG. The reduction of processes for Rate 1/8 decreases the computational complexity. The processes of the conversion for each parameters will be covered in detail in the following subsections.

3.2. LSP conversion

In the proposed transcoding algorithm, the LSP of SMV is converted to that of G.723.1 via linear interpolation. Each codec uses a different weighting window of different length in the LPC analysis. The interpolation coefficients are selected by considering these differences. Fig. 2 shows the weighting window used in each codec and the linear interpolation used for transcoding LSP of SMV into that of G.732.1. The LSP conversion in Fig. 2 can be written mathematically by

$$LSP_K^G = \frac{2}{3}LSP_L^{SMV} + \frac{1}{3}LSP_{L+1}^{SMV} \qquad (1)$$

$$LSP_{K+1}^G = LSP_{L+2}^{SMV} \qquad (2)$$

where LSP_L^{SMV} and LSP_K^G are the LSP of SMV and G.723.1 respectively and L and K are the frame index. In the proposed algorithm, LPC analysis and the conversion of LPC into LSP are not required, while they are required in the tandem transcoding. Because of this, the computational complexity is reduced. The absence of a lookahead, which is essential for LPC analysis, also reduces the algorithmic delay.

3.3. Pitch conversion

Fig. 3 shows the procedure of the proposed pitch conversion algorithm. For every two subframes of G.723.1, the open-loop pitch period is linearly predicted from the previously estimated open-loop pitch periods. The predicted pitch period is compared with the closed-loop pitch period of SMV. If the difference between both pitch periods is less than the threshold value, the closed-loop pitch searched by SMV is determined as the open-loop pitch period of G.723.1. Otherwise, the open-loop pitch search of G.723.1 is performed to find more precise pitch period. This pitch conversion using linear prediction has two advantages. First, it is possible to estimate the open-loop
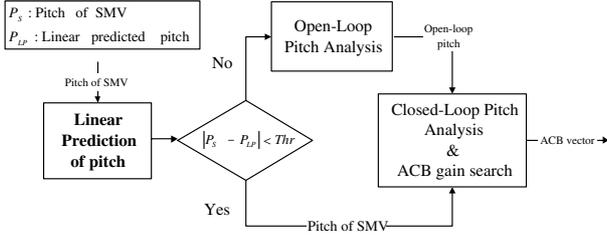
Figure 3: *Block diagram of pitch conversion (SMV → G.723.1)*

Figure 4: *Block diagram of transcoding algorithm from G.723.1 to SMV*

Figure 5: *Pitch conversion via linear interpolation (G.723.1 → SMV)*

pitch period with much less computational complexity. While the original open-loop pitch search of G.723.1 needs to calculate the cross-correlation of the perceptual weighted speech and detect its maximum value, the proposed pitch conversion only needs a prediction of the pitch period and a comparison of the predicted pitch period and the decoded pitch period of SMV. Second, since the open-loop pitch is re-estimated when there exists a large difference between the pitch periods obtained by both speech coders, degradation of the speech quality can be avoided.

4. Transcoding from G.723.1 to SMV

4.1. General Description

Fig. 4 shows a block diagram of the proposed transcoding algorithm in the case of the transcoding from G.723.1 to SMV. Overall structure in this case is similar to that of the transcoding from SMV to G.723.1, however the details of each processing are different. In addition, a simplified rate selection algorithm is added to the structure, since SMV needs the selection of the encoding rate for every frame. The necessary parameters are LSP, pitch period for ACB, ACB gain and FCB gain. In the parameter decoding, the necessary parameters are decoded from the input bit-stream of G.723.1 and the speech parameters for the simplified rate selection are calculated using the decoded parameters. After LSP conversion and open-loop pitch conversion, the proposed transcoding algorithm classifies the frame and selects the rate and type of the frame using the simplified rate selection algorithm. Finally, based on the selected rate and type, the excitation for the synthesis filter is estimated. For Rate 1 and Rate 1/2, the open-loop pitch determined in pitch conversion is used to search ACB. The FCB is searched as in the encoder of SMV. For Rate 1/4 and Rate 1/8, the excitation signal is generated by random number generator such as SMV.

4.2. LSP and Pitch conversion

In this case, the LSP conversion adopts the linear interpolation similar to the case of the transcoding from SMV to G.723.1. The linear interpolation in this LSP conversion is written by

$$LSP_L^{SMV} = \frac{7}{24}LSP_{K-1}^G + \frac{17}{24}LSP_K^G \qquad (3)$$

$$LSP_{L+1}^{SMV} = \frac{2}{3}LSP_K^G + \frac{1}{3}LSP_{K+1}^G \qquad (4)$$

$$LSP_{L+2}^{SMV} = LSP_{K+2}^G \qquad (5)$$

In the pitch conversion, the algorithm proposed in the transcoding from SMV to G.723.1 is not adopted. There are just a linear interpolation of pitch periods for matching the subframe of G.723.1 with the half frame of SMV. The adopted linear interpolation is shown in Fig. 5 where $P_K^G[i]$ and $P_L^S[i]$ are the pitch period of G.723.1 and SMV, respectively, L and K are the frame index and i is the subframe index.
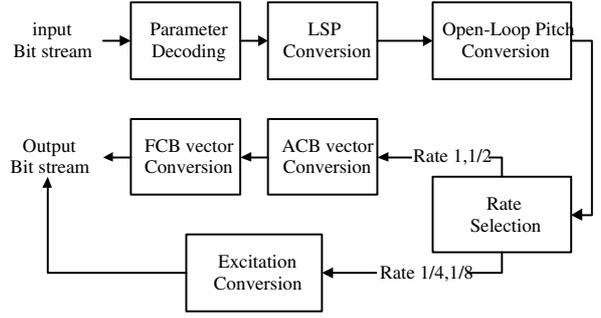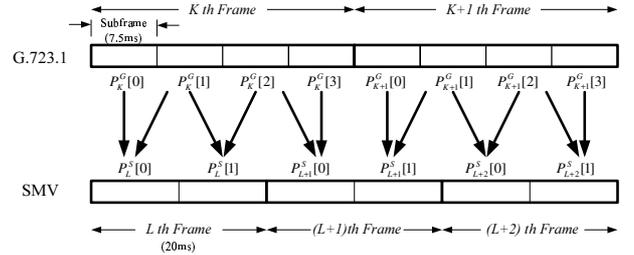
4.3. Rate selection

The SMV has to classify the input speech frame as one of 6 possible classes and choose appropriate rate and type. But those processes need various speech parameters that are calculated from the original input speech. To calculate the speech parameters from not the input speech but the decoded parameters of G.723.1, we proposed a simplified rate selection algorithm. In the proposed algorithm, the number of frame classes is reduced to 5. The classification of the reduced set of classes is shown in Fig. 6. Based on the extensive simulations, it is known that ACB gain is closely related to the voice activity of the speech. Thus, by comparing the smoothed ACB gain to the threshold value, the frame can be classified as silence, unvoiced or voiced. In addition, it is also known that the variance of pitch periods is large in the silence and noisy frame while it is small in the voiced frame. Therefore the variance of pitch period is considered for the determination of speech and silence too. FCB gain can also be exploited for the classification of the speech. FCB gain shows similar tendency to the voice activity of the speech, however it is easily influenced by the noise. Therefore, FCB gain is regarded with Noise-to Signal Ratio (NSR) which is calculated by the algorithm realized in SMV. Onset is determined in the case that there is a transition from unvoiced speech to voiced speech. For the division of stationary and nonstationary voiced, all ACB gains of the current frame are used. If those are bigger than threshold value, the frame is classified as stationary voiced. After the frame classification, the encoding rate is selected. The process is similar to that of SMV. Some speech parameters that don't need LP analysis and open-loop pitch analysis are computed and used.

5. Evaluation results

This section provides various evaluation results of the proposed transcoding algorithm. In the evaluation, utterances of 60 seconds-long spoken by both Korean male and female speaker
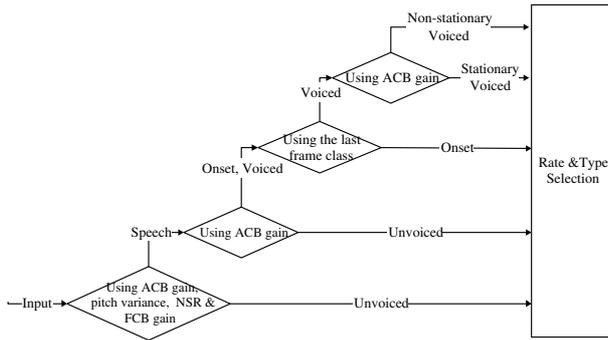
Figure 6: *Block diagram of simplified classification algorithm*

Table 1: *Computational complexity*

Speaker	Mode	SMV → G.723.1		G.723.1 → SMV	
		Tandem	Proposed	Tandem	Proposed
Male	0	19.95	12.59	29.46	23.30
	1	20.04	12.42	27.69	21.90
	2	20.24	12.32	27.44	21.53
	3	20.26	12.30	27.52	21.52
Female	0	20.01	13.95	30.22	24.97
	1	20.20	13.80	29.47	22.81
	2	20.30	13.79	28.47	22.65
	3	20.30	13.77	29.24	22.67

Table 2: *Speech Quality*

Speaker	Mode	SMV → G.723.1		G.723.1 → SMV	
		Tandem	Proposed	Tandem	Proposed
Male	0	3.237	3.239	3.303	3.244
	1	3.180	3.174	3.204	3.166
	2	3.098	3.135	3.112	3.050
	3	3.026	3.091	3.095	2.982
Female	0	3.066	3.023	3.081	3.102
	1	2.996	2.997	2.959	3.069
	2	2.942	2.960	2.904	2.863
	3	2.915	2.973	2.957	2.817

are used. For each simulation, the performance of the proposed transcoding algorithm is compared with that of the tandem transcoding algorithm.

5.1. Computational Complexity

To evaluate the computational complexity of the proposed transcoding algorithm, the weighted million operations per second (WMOPS) is calculated and compared to that of the tandem transcoding algorithm. Table 1 shows the WMOPS values of the tandem and proposed transcoding algorithms. As shown in Table 1, the computational complexity of the proposed algorithm is about 20-35% lower than that of the tandem transcoding algorithm.

5.2. Speech Quality

To measure the speech quality of transcoding algorithm, we used the perceptual evaluation of speech quality (PESQ)[7]. PESQ values of the tandem and proposed transcoding algorithm are compared in Table 2. The results in Table 2 indicates that the overall quality of the proposed transcoding algorithm is slightly better than or equal to that of the tandem transcoding algorithm.

5.3. Delay

In the evaluation of the delay, we consider only algorithmic and processing delays except the transmission delay. In the transcoding from SMV to G.723.1, the look-ahead for LPC analysis is not required and it decreases the algorithmic delay. But the overall delay in the opposite direction, is not reduced by the look-ahead. It is because the information of speech interval corresponding to the look-ahead is used for other processes such as rate selection in the SMV. In all directions, the processing delay is reduced by omitting unnecessary processes.

6. Conclusion

In this paper, we proposed a novel transcoding algorithm for SMV and G.723.1 speech coders via direct parameter transformation. Evaluation results show that the proposed algorithm can achieve equivalent speech quality to that obtained by tandem transcoding with reduced computational complexity and delay.

7. 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] 3GPP2 Spec. "Selectable Mode Vocoder Service Option for Wideband Spread Spectrum Communication Systems", 3GPP2-C.S0030-0 v2.0, Dec. 2001.

[3] Yang Gao, A. Benyassine, J. Thyssen, Huan-yu Su, E. Shlomot, "EX-CELP : A Speech Coding Paradig", In *Proc. ICASSP 2001*, vol. 2, pp. 689-692, 2001.

[4] Hong-Goo Kang, Hong-Kook Kim, R.V. Cox, "Improving transcoding capability of speech coders in clean and frame erasured channel environments," In *Proc. IEEE Workshop on Speech Coding, 2000*, pp. 78–80, Jan., 2000.

[5] Sung Wan Yoon, Sung Kyo Jung, Young Cheol Park, and Dae Hee Youn, "An efficient transcoding algorithm for G.723.1 and G.729A speech coders", In *Proc. Eurospeech 2001*, vol. 4, pp. 2499-2502, 2001.

[6] Sunil Lee, Seongho Seo, Dalwon Jang, Chang D. Yoo, "A novel transcoding algorithm for AMR and EVRC speech coders via direct parameter transformation," In. *Proc. ICASSP 2003*, 2003 (ACCEPTED).

[7] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," 2000.