IMPROVEMENT ISSUES ON TRANSCODING ALGORITHMS : FOR THE FLEXIBLE USAGE TO THE VARIOUS PAIRS OF SPEECH CODEC

Jin-Kyu Choi, Chang-Heon Lee, Hong-Goo Kang*, Young-Cheol Park** and Dae Hee Youn**

MCSP Lab., Yonsei University / LG Electronics Inc., Korea
*Center for Information Technology, MCSP Lab., Yonsei University, Korea
**Division of Information Technology, Yonsei University, Wonju-city, Korea
jinkyu@mcsp.yonsei.ac.kr

ABSTRACT

This paper describes important issues on transcoding between different speech codecs by considering the paradigms of source and target coders. Conventional transcoding algorithms on LSP, pitch and adaptive/fixed codebook conversion are refined with regard to the structure of coders. In addition, a new perceptual weighting filter that plays a role in post-filter and perceptual weighting filter together is proposed to further improve the performance. The performance of the proposed algorithms is verified in step-by-step manner with the examples of transcoding between AMR, G.723.1 and G.729. By employing the proposed algorithm to the transcoders, the complexity is reduced by about 20%$\sim$76.88% and quality is also improved compared to conventional approaches.

1. INTRODUCTION

In applications requiring interoperability between the different networks such as wireless and voice over IP (VoIP), transcoding is a good choice due to its lower complexity, delay and quality degradation [1][2]. Its structure is a cascaded connection of a decoder of transmitter side and an encoder of receiver side, but converts bit-streams of source coder to those of target coder as much as possible. Transcoder consists of several routines such as LSP conversion, pitch estimation, adaptive/fixed codebook search and new weighting filter which reduce the complexity without quality degradation. The detailed transcoding algorithms should be varied depending on the characteristics of codecs such as frame size, pitch and adaptive/fixed codebook structure.

This paper proposes several algorithms that can be used in general transcoding. In detail, the algorithm is described with the examples of transcoders between AMR, G.723.1 and G.729A that are the most popular speech coders in wireless and VoIP applications [3][4][5]. Since all these coders have analysis-by-synthesis scheme and their transmitted information are similar, transcoding can be applied reasonably well. However, since the analysis frame size for each coding is different, 20ms to AMR, 30ms to G.723.1 and 10ms to G.729A, cares must be taken to successfully deploy the systems. In addition, the pitch intervals for adaptive codebook are different, we need special processing not to degrade the quality. Standard speech coders have post filters to hide the quantization noise that occurs during the encoding processes. Though it is very efficient for a single encoding case, it might not be necessary for transcoding because the same effect can be obtained by controlling the coefficients of the perceptual weighting filter in target encoding. This topic has not been considered before, but it turns out that it is very effective in terms of quality and complexity. The proposed algorithms can be easily extended to other transcoding cases.

This work was supported by Electronics and Telecommunications Research Institute (ETRI).

2. TRANSCODING BETWEEN AMR, G.723.1 AND G.729A

Since the three coders that we consider in this paper have differences in frame size, the structure of adaptive/fixed codebook and so on, many problems should be considered during transcoding. In this section, several schemes considering those differences are described in detail. As a matter of convenience, in tables, “A”, “G1” and “G2” mean AMR, G.723.1 and G.729A, respectively. bit-rates is shown in this paper.

2.1. LSP Conversion

Typically, spectral coefficients such as LSP parameters are obtained by inverse quantization process of the transmitter side coder and by encoding the coefficients with the quantization table of the target encoder [1][6]. Thus, the processing such as LP analysis is not needed in transcoding. If subframes of source and target coders are synchronous, then the LSP parameters of source coder can be used as those of target coder. But, if not, additional processing such as interpolation or merging procedure is required to generate suitable LSP parameters. Generally, in case of coders having different frame sizes, the LSP sets of target coder should be generated by interpolation whose coefficients are determined from geometrical distance. If the frame size of source coder is small, then the interval of LSP interpolation is small.

In case of transcoder from G.729A to other coders, the additional processing is not required because the direct mapping is possible due to frame synchronization and sufficient LSP information. But, in other transcoding cases, the LSP information is not enough to convert the parameters of source coder to those of target coder. Consequently, the additional processing such as interpolation is needed. During this process, conventional LSP conversion scheme considers only geometrical distance between the

Table 1. PESQ of LSP conversion

AMR(kbps)		5.15	7.4	7.95	12.2
G1(5.3) → A	**Tandem**	3.079	3.235	3.250	3.341
	LSP	3.063	3.249	3.300	3.376
G1(6.3) → A	**Tandem**	3.152	3.328	3.342	3.460
	LSP	3.110	3.324	3.343	3.473
A → G1(5.3)	**Tandem**	3.077	3.225	3.232	3.346
	LSP	3.160	3.279	3.300	3.347
A → G1(6.3)	**Tandem**	3.148	3.329	3.337	3.469
	LSP	3.232	3.389	3.411	3.469
A → G2	**Tandem**	3.189	3.387	3.395	3.519
	LSP	3.286	3.440	3.461	3.492
G2 → A	**Tandem**	3.160	3.345	3.359	3.463
	LSP	3.184	3.432	3.445	3.542

given LSP and the target LSP. But, considering the look-ahead delay of target coder, we can further improve the performance. This means that the look-ahead delay is ignored in transcoding algorithms but this delay must be considered in setting the coefficients of interpolation. The proposed LSP conversion scheme considers not only geometrical but also the look-ahead delay by setting the factor with virtual distance of look-ahead delay of target coder.

In LSP conversion, we must utilize the given information as much as possible. As examples, in case of transcoder from G.729A to AMR, G.729A and AMR have equal subframe sizes though frame sizes are different. Thus, it is simple that the LSP sets of 2nd G.729A frame are mapped to those of 1st AMR frame. To further improve quality, both LSP sets of 1st and 2nd frames of G.729A are used in LSP interpolation to minimize the mean of the spectral distortion of interpolated LSP sets [6].

Table 1 shows the performance of this scheme in transcoding pairs of AMR, G.723.1 and G.729A when only LSP is converted. From the results of transcoder between AMR and G.723.1, we know that the performance of the transcoder having small frame size of source coder is better than opposite case. In case of transcoder between AMR and G.729A, the quality is also improved with lower complexity due to LSP conversion.

2.2. Pitch Estimation

The pitch is an important parameter in speech coding and the processing for estimating them per subframe requires high complexity. In transcoding algorithm, this complexity can be removed or drastically reduced without quality degradation by using a direct mapping or a pitch smoothing method [1].

If the subframe sizes of source and target coders are exactly equal such as transcoding between AMR and G.729A, the pitches of source coder can be used as those of target coder without any additional processing. However, if the subframe sizes are different such as transcoding between AMR and G.723.1, then additional processing is needed to generate the pitches of target coder. In these cases, the pitch smoothing method using smoothing effect is one of the good approaches. This scheme reduces the complexity with a little quality degradation [1]. However, this method should have high complexity almost equal to that of full search in unvoiced frames, because its complexity depends on the difference of the pitch intervals from source and target coders.

We propose a fast pitch estimation algorithm to further reduce complexity, while improving quality. The proposed algorithm includes some of adjacent samples near the closed-loop pitch of source coder to search the open-loop pitch of target coder. The proposed scheme is denoted by

$$C_{OL}(j) = \frac{\left(\sum_{n=0}^{N} s_w[n]\cdot s_w[n-j]\right)^2}{\sum_{n=0}^{N} s_w[n-j]\cdot s_w[n-j]}\ , P_{min} \le j \le P_{max} \tag{1}$$

where, the index j which maximizes the cross-correlation, $C_{OL}(j)$, is selected as the open-loop pitch for the appropriate subframes of N samples. s_w is the perceptually weighted speech. P_{min} and P_{max} are the bounds of the search range of the proposed algorithm. This bound is selected by considering frame synchronization of transcoder and characteristics of each coders.

Fig.1 shows PESQ [7] by varying the search bound of the proposed algorithm. In case of transcoding from G.723.1 to AMR, compatible quality can be obtained by using just 3 closest samples for the pitch conversion. Conversely, the proposed algorithm requires longer interval of the bound in case of the transcoding from AMR to G.723.1. We choose 7 samples considering the quality and complexity aspects. Moreover, from the result of transcoder between AMR and G.729A, we know that the pitches of source coder can be used as those of target coder without any quality degradation because two coders have equal subframe sizes [3][5]. Thus, in case of transcoding between AMR and G.729A, the direct mapping method is used in this paper.

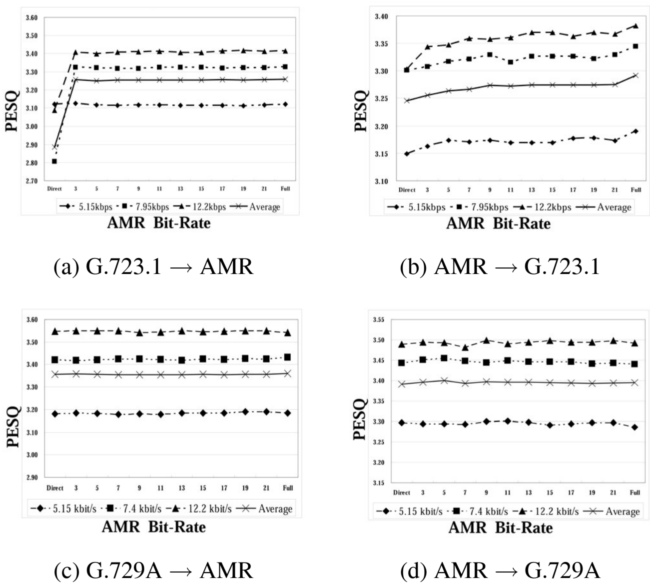
Table 2 and 3 show the PESQ [7] and WMOPS [8] of transcoder between G.723.1 and AMR, respectively. The proposed algorithm reduces over 90% complexity of this routine without quality degradation. In case of transcoder between AMR and G.729A, the complexity of this routine is completely removed due to direct mapping. From these results, we can say that the search bound of pitch interval can be defined by considering not only the range search but also the estimation processes of source and target coder. As examples, the closed-loop pitch of G.723.1 is more accurate than that of AMR due to using more samples(large window) for estimating the pitch [9]. Thus, the bound of the proposed algorithm in case of transcoding form G.723.1 to AMR can be smaller than that in opposite case. This proposed algorithm can be easily extended to other transcoding cases.

Table 2. PESQ of fast pitch estimation

AMR(kbps)		5.15	7.4	10.2	12.2
G1(5.3) → A	**Full Search**	3.120	3.306	3.397	3.431
	Proposed	3.125	3.301	3.399	3.424
G1(6.3) → A	**Full Search**	3.169	3.384	3.504	3.548
	Proposed	3.161	3.376	3.499	3.539
A → G1(5.3)	**Full Search**	3.176	3.308	3.363	3.387
	Proposed	3.165	3.277	3.334	3.350
A → G1(6.3)	**Full Search**	3.251	3.410	3.478	3.507
	Proposed	3.240	3.386	3.458	3.492
A → G2	**Full Search**	3.286	3.440	3.465	3.492
	Direct	3.297	3.443	3.453	3.489
G2 → A	**Full Search**	3.184	3432	3.488	3.542
	Direct	3.181	3.420	3.495	3.547

2.3. New Perceptual weighting Filter

A perceptual weighting filter de-emphasizes the formant regions and a post filter compensates for this effect. In the conven-



(a) G.723.1 → AMR (b) AMR → G.723.1

(c) G.729A → AMR (d) AMR → G.729A

Fig. 1. Search bound of pitch estimation

Table 3. WMOPS of fast pitch estimation

G.723.1(kbps)		**5.3**	**6.3**
A → G1	**Full Search**	1.402	1.545
	Proposed	0.066	0.066
	Reduction	95.3%	95.73%

AMR(kbps)		**5.15**	**7.4**	**10.2**	**12.2**
G1 → A	**Full Search**	1.133	1.231	1.439	1.251
	Proposed	0.093	0.120	0.120	1.120
	Reduction	91.8%	90.3%	91.7%	90.4%

tional transcoder, these two filters should be operated in cascade, and each filter emphasizes or de-emphasizes formant regions. During the procedures, undesirable phase distortion may be occurred because the post and perceptual weighting filter have non-linear phase characteristics. In addition, the computational loads of those two filters are somewhat high, especially for the processing of post filter[3][4]. To solve those problems, a new filter that can replace two filters by one is designed in this section.

In general transcoder, the same or little different LPCs are used in the post and perceptual weighting filter because LPCs are directly connected [1][6]. Thus it is possible to assume that the formant locations of both filters are same. The cascaded filter of post and perceptual weighting filter is written by

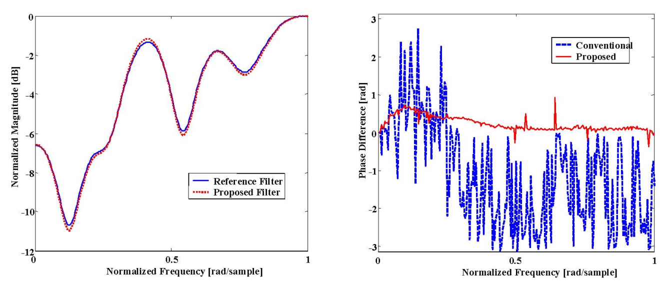
$$H_R(z) = \frac{\prod_{i=1}^{p}(1 - a_i e^{j\omega_i} z^{-1}) \prod_{i=1}^{p}(1 - c_i e^{j\omega_i} z^{-1})}{\prod_{i=1}^{p}(1 - b_i e^{j\omega_i} z^{-1}) \prod_{i=1}^{p}(1 - d_i e^{j\omega_i} z^{-1})}, \quad (2)$$

where p is LPC order(10th order is used generally), a_i, b_i, c_i and d_i are real numbers.

As shown in Eq.(2), the poles and zeros of the post filter and the perceptual weighting filter are located at the same or little different angles because same or interpolated LSPs are used in both filters. Thus, it can be assumed that they are same in both sides. It is also possible to design a new filter that has the same magnitude response as them.

The proposed algorithm is implemented by same structure of the perceptual weighting filter. The magnitude response of new filter is controlled by the weighting factors. The factors can be determined to have minimum mean squared error by iterative algorithm (MMSE), where the reference filter is the cascaded filter of post and perceptual weighting filter and the error is defined as the difference in the magnitude responses of the reference filter and proposed filter.

Fig.2 show that not only the magnitude response of the new filter is almost equal to that of the reference filter, but also the phase distortion is drastically degraded. By using the proposed filter, the complexity of the post filter is completely reduced and quality is significantly improved. This improvement sufficiently compensates for the quality degradation due to fast algorithms such as fast open-loop pitch estimation and fast adaptive codebook search. By applying this scheme to transcoder between AMR and G.723.1, this result will be shown in next section.



(a) Magnitude response (b) Phase distortion

Fig. 2. New weighting filter

2.4. Adaptive/Fixed codebook

Adaptive codebook parameters of source coder can be used as those of target coder when the structures of two coders are equal or similar such as transcoding between AMR and G.729A having equal structure but different resolutions. In case of transcoder between AMR and G.723.1, this scheme can not be employed due to different structures, fractional pitch lag to AMR and fifth order predictor to G.723.1 [3][4]. Fixed codebook conversion without additional procedure is possible only if the structures of two coders are exactly same such as the transcoding of AMR(7.4 and 7.95 kbps) and G.729A. More researches are required for those problems.

3. PERFORMANCE EVALUATION

In this section, we show the performance of the proposed algorithm being applied to the pairs of G.723.1 and AMR, and analyze the associated modules. PESQ [7] and WMOPS [8] with 96 speech samples (NTT-AT Korean data) are used as objective quality and complexity measurement, respectively. Transcoder from AMR to G.723.1 used in the experiments employs the proposed and fast ACB search schemes [1]. Fig.3 shows that the quality is improved by employing LSP conversion and new weighting filter but is somewhat degraded by employing modified pitch estimation and fast ACB search scheme. In terms of final quality, the proposed transcoding algorithm is better than tandem case. Transcoder between AMR and G.729A employs several proposed schemes such as LSP conversion, modified pitch estimation and adaptive/fixed codebook mapping scheme. Adaptive/fixed codebook (ACB/FCB) mapping scheme can not be used in some cases because of different structures associated to bit-rates. Fig.4 shows the performance of the transcoder from G.729A to AMR. By using the proposed algorithms, the performance is also improved in this transcoding case compared to tandem.

Table 4 shows the reduction ratio of complexity associated to the pairs of transcoder between AMR, G.723.1 and G.729A. The proposed algorithm reduces the complexity by about 20%∼76.88%. The maximum reduction is achieved when the direct mapping is possible in many routines because the characteristics of source and target coder are almost equal.

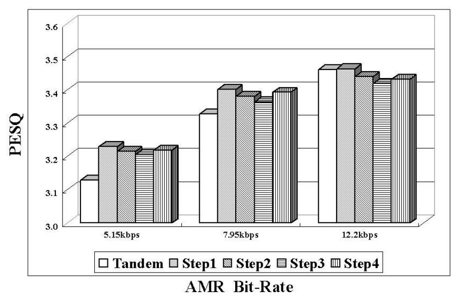


Fig. 3. PESQ of Transcoder (AMR → G.723.1) (Step1: LSP conversion, Step2: Step1 + Modified Pitch Estimation, Step3: Step2 + Fast ACB Search, Step4: Step3 + New Weighting Filter)

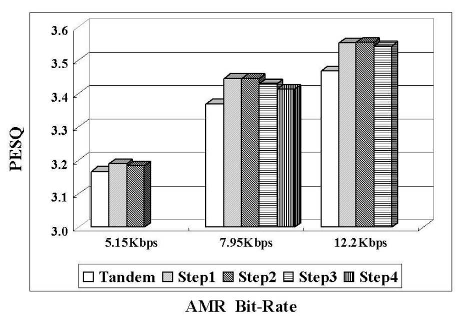


Fig. 4. PESQ of Transcoder (G.729A → AMR) (Step1: LSP conversion, Step2: Step1 + Modified pitch estimation, Step3: Step2 + ACB mapping, Step4: Step3 + FCB mapping)

4. CONCLUSION

This paper proposed several transcoding algorithms by splitting them with LSP conversion, modified pitch estimation and new weighting filter. To successfully deploy the transcoding algorithms, we should carefully consider the paradigms of standard speech coders such as frame size, pitch interval, coefficients for perceptual weighting and so on. By applying the proposed algorithm to the pairs of AMR, G.723.1 and G.729A, we could improve the quality

Table 4. Complexity of Transcoder

Method \ AMR Bit-Rate(kbps)		**WMOPS 5.15**	**7.95**	**12.2**
A→G1(5.3)	Tandem	42.750	42.801	42.819
	Transcoding	18.954	19.005	19.023
	Reduction(%)	55.663	55.597	55.573
G1(5.3)→A	Tandem	28.334	38.498	38.360
	Transcoding	19.566	29.517	27.171
	Reduction(%)	25.701	19.127	25.272
A→G2	Tandem	9.842	9.860	9.867
	Transcoding	5.703	2.805	6.966
	Reduction(%)	42.06	71.55	29.40
G2→A	Tandem	10.395	13.788	13.449
	Transcoding	7.105	3.188	8.230
	Reduction(%)	32.52	76.88	38.81

and also reduce the complexity by about 20%∼76.88% compared to tandem. The proposed algorithm can be easily applied to general transcoding algorithms if we consider their own characteristics.

As many different speech coders are commercialized, combining different pairs of transcoding algorithms into one system with an efficient way should be a critical issue.

5. REFERENCES

[1] S. W. Yoon, J. K. Choi, H. G. Kang, and D. H. Youn, “Transcoding Algorithm for G.723.1 and AMR Speech Coders: for Interoperability between VoIP and Mobile Networks” in *Proc. Eurospeech 2003*, pp.1101∼1104, Sep., 2003.

[2] Sunil Lee, S. Seo, D. Jang and C. D. Yoo “A novel transcoding algorithm for AMR and EVRC speech codecs via direct parameter transformation” in *ICASSP 2003*, pp.II-177∼II-180, 2003.

[3] 3GPP TS 26.090 V5.0.0, “AMR speech codec; Transcoding functions,” Jun., 2002.

[4] ITU-T Rec. G.723.1, “Dual-rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s,” May 1996.

[5] ITU-T Rec. G.729 Annex A, “Reduced complexity 8 kbit/s CS-ACELP Speech codec,” 1996.

[6] Hong-Goo Kang, Hong-Kook Kim and Richard V. Cox, “Improving the Transcoding Capability of Speech Coders” in *IEEE Transaction on Multimedia*, VOL.5, NO. 1, Mar., 2003.

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

[8] ITU-T Draft Rec P.191, “Software tools for speech and audio coding standardization,” Nov., 2000.

[9] W.B. Kelijn, *SPEECH CODING AND SYNTHESIS*, Elsevier Science B.V., 1995.