A NOVEL TRANSCODING ALGORITHM FOR AMR AND EVRC SPEECH CODECS VIA DIRECT PARAMETER TRANSFORMATION

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

Multimedia Processing Lab., Dept. of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology
373-1 Guseong-dong, Yuseong-gu, Daejeon, Republic of Korea, 305-701
cowboysun@mail.kaist.ac.kr

ABSTRACT

In this paper, a novel transcoding algorithm for the Adaptive Multi Rate (AMR) codec and the Enhanced Variable Rate Codec (EVRC) is proposed. In contrast to the conventional tandem transcoding algorithm, the proposed algorithm transcodes the parameters of one codec to the other without synthesizing the speech. The proposed algorithm decodes the parameters of source codec from the input bitstream, and based on frame classification and mode decision, it appropriately transforms the parameters of source codec to that of the target codec in the parametric domain. Finally, the transformed parameters are encoded into a bitstream that is decodable by the target codec. The parameters transcoded by the proposed algorithm are line-spectral pair (LSP), pitch delay, fixed codebook vector, codebook gains, and frame energy. Evaluation results show that while reducing both the computational complexity and delay by 50%, the proposed algorithm produces speech quality equivalent to that of produced by the tandem transcoding algorithm. The general idea is not restricted to the AMR and EVRC but is applicable to various other code-excited linear prediction (CELP) based codecs.

1. INTRODUCTION

Today, there exists a wide variety of wire and wireless communication networks in which different speech coding standards are adopted. The development of an efficient transcoding algorithm for different speech codecs is an important issue for the integration and interoperability of different networks. Generally, the simplest way to solve the compatibility problem between two different speech codecs is by the tandem transcoding algorithm : generate speech signal using a decoder of one speech codec and then re-encode the signal by the other speech codec. But this approach results in loss in speech quality, increase in complexity and delay as a result of additional decoding and encoding process that the signal has to go through in the transcoding process. These problems can be alleviated by direct parameter transformation in which the parameters between the two codecs are transcoded without generating the speech.

This paper proposes a novel transcoding algorithm for the Adaptive Multi Rate (AMR)[1] codec and the Enhanced Variable Rate Codec (EVRC)[2] by direct parameter transformation. The algorithm takes advantage of the fact that both codecs are based on the code-excited linear prediction (CELP)[3] paradigm and share similar parameter set-LSP, pitch delay, fixed codebook vector, codebook gains and frame energy. Though there are similarities, the two codec are very different in many respects. The AMR codec has been chosen by the Third Generation Partnership Project (3GPP) as the mandatory codec for the third generation (3G) cellular systems. It supports 8 encoding modes with bit rates between 4.75 and 12.2 kbit/s. On the other hand, the EVRC is a speech coding standard for the 2nd generation CDMA system. It is also a multi-rate codec that supports 3 encoding modes — rate 1/8(0.8 kbit/s), 1/2(4.0 kbit/s) and 1(8.55 kbit/s) — and is based on relaxation CELP (RCELP)[5] algorithm.

The proposed algorithm decodes the parameters of one codec from the input bit-stream, and based on frame classification and mode decision, it appropriately transforms the parameters of one to that of the other in the parametric domain. Finally, the transformed parameters are encoded into a bit-stream that is decodable by the other codec. The parameters transcoded by the proposed algorithm are line-spectral pair(LSP), pitch delay, fixed codebook vector, codebook gains, and frame energy.

The rest of the paper is organized as follows. In Section 2 of this paper, we describe the proposed transcoding algorithm via direct parameter transformation in detail. Section 3 provides the results of the various performance evaluation on the proposed algorithm. Finally, Section 4 concludes the paper.

2. PROPOSED TRANSCODING ALGORITHM

2.1. General Description

The proposed algorithm consists of three modules: 1)the parameter decode, 2) frame classification and mode deci-

This work was supported by grant No. R01-2000-000-00259-0(2002) from the Korea Science & Engineering Foundation.

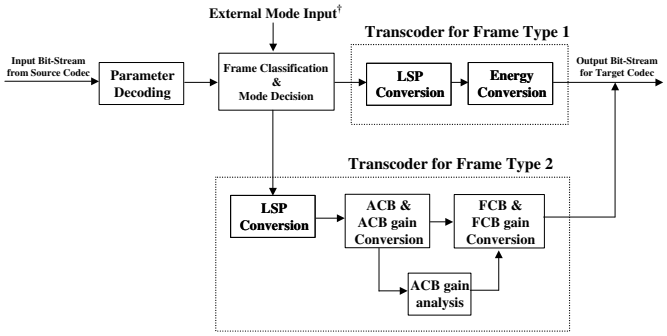

Fig. 1. Simplified block diagram of the proposed transcoding algorithm († exists only in the transcoding from the EVRC to AMR)

sion, and 3)parameter transformation and encoding modules. Figure 1 shows a simplified block diagram of the proposed transcoding algorithm.

In the parameter decode module, the parameters transmitted from the source codec are decoded. The parameters to be decoded are LSP, pitch delay, fixed codebook vector, codebook gains, and frame energy. The information on the encoding mode can also be obtained from this module.

In the frame classification module, based on the values of decoded parameters, the input speech frame is classified either as *frame type 1* or *frame type 2*. Silence or background noise are classified as *frame type 1*. Other frames which contain active speech are classified as *frame type 2*. Since the silence or background noise is encoded and transmitted with special low bit rate mode(DTX mode of the AMR and Rate 1/8 mode of the EVRC), the classification decision can be made based on the mode information obtained in the parameter decoding module.

Classified input frames are transcoded according to their type. Since the parameters to be transcoded are different for each frame type, there are two transcoders for each frame type. While the transcoder for *frame type 1* converts only two parameters—LSP and frame energy, the transcoder for *frame type 2* converts the LSP, pitch delay, fixed codebook vector, and codebook gains. The processes of the transcoding for each parameter will be covered in detail in the following subsections.

2.2. Conversion of LSP

In the proposed algorithm, the LSP parameter is converted by simple linear interpolation. Although the frame lengths of both codecs are same (20ms which corresponds to 160 speech samples), the shapes of the analysis window and the lookahead periods are different. Considering these differences, the decoded LSP of the m^{th} and $(m-1)^{th}$ frames of the source codec, $\Omega_A^{(m)}$, $\Omega_A^{(m-1)}$, are linearly combined to give the m^{th} frame LSP of the target codec, $\Omega_B^{(m)}$, by

$$\Omega_B^{(m)} = \mu\Omega_A^{(m)} + (1-\mu)\Omega_A^{(m-1)} \quad (1)$$

where μ is the weighting constant. The value of μ is empirically chosen by minimizing the average spectral distortion measure(ASDM)[6] between the original and the converted LPC spectra. Based on extensive simulation, $\mu = 0.84$ for the transcoding of AMR to EVRC and $\mu = 0.96$ for the transcoding of EVRC to AMR were chosen and used in this paper.

2.3. Conversion of Pitch Delay

The pitch delay can not be directly transformed as in the case of LSP since it would seriously degrade the speech quality. Thus additional pitch search is necessary in the transcoding process. The AMR obtains and transmits the pitch delay to a fractional resolution every subframe whereas the EVRC only one integer pitch delay every frame since it is based on the RCELP algorithm. For speech frames declared *type 2*, the absolute error between the pitch delays of each codec is less than 10 speech samples [7]. Thus the pitch delay calculated from one codec can be used to restrict the search range of the other codec.

In the transcoding from AMR to EVRC, first, the average pitch delay T_0 is calculated by taking the mean of four pitch delays of each subframe. Secondly, T_0 is compared with the converted pitch delay of the previous frame τ_{-1}. If the value of T_0 is not in the range of $[0.8\tau_{-1}, 1.2\tau_{-1}]$, a new pitch delay for the current frame of the EVRC τ_0 is obtained via full search. If T_0 is in the range, τ_0 is searched near T_0 as follows. The value of D_{max} which maximizes $R(D)$ given below

$$R(D) = \sum_{n=0}^{159-D} \varepsilon[n+n_0]\varepsilon[n+n_0+D]$$
$$max\{20, T_0-5\} \le D \le min\{120, T_0+5\} \quad (2)$$

where $\varepsilon[n]$ is the excitation signal calculated using the decoded parameters is obtained for $n_0 = 80$ and 160. Finally, converted pitch delay τ_0 is decided between the two D_{max} values using the decision rule used in the EVRC.

In the transcoding from the EVRC to AMR, first, the decoded pitch delay τ_0 of EVRC is compared with the converted pitch delay (AMR) of the previous frame T_{-1}. If τ_0 is in the range $[0.8T_{-1}, 1.2T_{-1}]$, τ_0 is directly used as an estimate of the open-loop integer pitch delay for the closed-loop pitch delay search of the AMR. Otherwise, new pitch delay T_0 is obtained via full search.

2.4. Conversion of Fixed Codebook Vector

Although the structures of the fixed codebooks of the AMR and EVRC are based on the same algebraic CELP (ACELP) [4] algorithm, there is little correlation between the fixed codebook vectors of each. Considering fixed codebook vectors as a kind of time series sequence, it can be empirically verified that the fixed codebook vectors searched by each codec are statistically independent of one another [7].

For this reason, it is impossible to directly convert the fixed codebook vector from one to the other.

The computational load of fixed codebook search accounts for a large portion(about 50%) of the total computation load required for tandem transcoding and thus fixed codebook search must be reduced at all cost. In this paper, a fast search algorithm is applied to reduce the computational load without the loss of perceptual speech quality.

In the fast search algorithm, the possible positions of the non-zero pulses are limited to those positions of large absolute value of the backward filtered target signal $\mathbf{d} = H^t\mathbf{x}_f$ where the vector $\mathbf{x}_f$ is the target signal for fixed codebook search and H is the lower triangular Toepliz convolution matrix. The target signal is chosen as a reference signal since the positions where this signal has a large absolute value are where it is more probable that a non-zero pulse exist than the others.

The number of possible position is determined by the adaptive codebook gain. By definition, the adaptive codebook gain $g_p =< \mathbf{x}_a \cdot H\mathbf{v} > / < H\mathbf{v} \cdot H\mathbf{v} >$ indicates how well the adaptive codebook vector models the excitation signal. The vector $\mathbf{x}_a$ is the target signal for the adaptive codeboook search and $\mathbf{v}$ is the adaptive codebook vector. The operation $< \cdot >$ stands for the inner product. For this reason, we use the adaptive codebook gain to control the number of combination of the pulse positions. Thus when the adaptive codebook gain is high, we put a high restriction on the number and vice versa when the gain is low. By applying this fast search algorithm, about 20–35% of the computational load for the fixed codebook search could be reduced.

2.5. Conversion of Codebook Gains and Frame Energy

The adaptive codebook gains g_p of AMR and EVRC are different for the same speech frame. Thus g_p of one codec can not be used for the other. Fortunately, g_p is calculated as a by-product in the pitch-delay transcoding process. The fixed codebook gain g_c can be converted by a simple linear interpolation as in the conversion of LSP. The converted g_c was nearly same to that calculated by the target codec.

The frame energy is calculated and transmitted only in the silence insertion descriptor(SID) frame and used to generate the comfort noise in the receiver. In the proposed algorithm, the frame energy is directly converted via simple adjustment of the magnitude.

3. PERFORMANCE EVALUATION

For the evaluation, we made a fixed point implementation of the proposed transcoding algorithm using the C programming language and 32 Korean sentences spoken by 4 male and 4 female speakers were used as input speech in all simulations. All sentences are 4 seconds-long, clean, and sampled at 8 kHz. No transmission error is assumed.

Table 1. Comparison of computational complexity with WMOPS

Algorithm	**AMR → EVRC**		**EVRC → AMR**	
	Male	Female	Male	Female
Tandem	13.62	14.29	13.20	13.22
Proposed	7.48	8.22	5.58	6.01
Reduction	45%	42%	58%	55%

3.1. Computational Complexity

We compared the computational complexity of the proposed algorithm to that of the conventional tandem transcoding algorithm by measuring the weighted million operations per second (WMOPS). As shown in Table 1, the average computational complexity of the proposed transcoding algorithm is about 42–58% lower than that of the tandem transcoding algorithm. The reduction of computational complexity can be attributed to the omission of additional LP analysis and adopting fast adaptive and fixed codebook search techniques.

3.2. Delay

Total delay of the speech communication system is calculated by summing the algorithmic, processing, and transmission delays. In this paper, however, we do not consider the transmission delay since it depends on the structure and status of the network. The total delays of the conventional tandem transcoding algorithm(D^{td}_{AB}, D^{td}_{BA}) and the proposed transcoding algorithm(D^{tl}_{AB}, D^{tl}_{BA}) are given by

$$D^{td}_{AB} = 35 + \alpha_A + \beta_A + \alpha_B + \beta_B \quad (3)$$

$$D^{tl}_{AB} = 25 + \alpha_A + P_{AB} + \beta_B \quad (4)$$

$$D^{td}_{BA} = 35 + \alpha_B + \beta_B + \alpha_A + \beta_A \quad (5)$$

$$D^{tl}_{BA} = 30 + \alpha_B + P_{BA} + \beta_A \quad (6)$$

where A is the source codec, B is the target codec, and AB is the transcoding from A to B, and BA is B to A, respectively. α_m and β_m ($m = A$ or B) are the processing delays of the encoders and decoders of the both speech codecs, while P_{AB} and P_{BA} are the processing delays of the proposed transcoding algorithm. As shown in the equations above, the delay of the proposed algorithm is at least 10 ms(AB) or 5 ms(BA) shorter than that of the tandem transcoding algorithm. The total delay is reduced since the proposed algorithm does not perform additional LP analysis, thus lookahead period(5 ms in AMR, 10 ms in EVRC) is no more required. Furthermore, the total delay of the proposed algorithm is much shorter than that of the tandem transcoding algorithm since $P_{AB} << \beta_A + \alpha_B$ and $P_{BA} << \beta_B + \alpha_A$. The results in Table 2 verify that the

Table 2. Comparison of processing delay

Algorithm	AMR → EVRC	EVRC → AMR
Tandem	1.42 sec	1.37 sec
Proposed	0.82 sec	0.75 sec
Reduction	43%	45%

Table 3. Comparison of PESQ score

	AMR → EVRC		EVRC → AMR	
Algorithm	Male	Female	Male	Female
Tandem	3.42	3.10	3.48	3.20
Proposed	3.44	3.09	3.47	3.11

processing delay of the proposed transcoding algorithm is much shorter than that of the conventional tandem transcoding algorithm. The simulation is performed on the PC platform of Pentium-IV 1.9GHz CPU, 512MB main memory, and Microsoft Windows XP operating system.

3.3. Objective Speech Quality Evaluation

We chose the perceptual evaluation of speech quality(PESQ) [8] as an objective speech quality assessment model. It is known that the absolute error between the PESQ and subjective scores is less than 0.25 MOS for 69.2% of the conditions and less than 0.5 MOS for 91.3% of the conditions. The average PESQ scores of the conventional tandem and proposed transcoding algorithms are compared in Table 3. The PESQ score of the proposed algorithm is nearly equivalent to that of the tandem algorithm. This indicates that the overall speech quality of the proposed transcoding algorithm is similar to that of the tandem algorithm.

3.4. Subjective Speech Quality Evaluation

We also performed an informal ABX preference listening test. The speech sentences mentioned above were transcoded by both the conventional tandem and the proposed transcoding algorithm. The subjects listened to each transcoded speech signals and determined whether the speech quality of one of them is better than the other or equivalent. The results in Table 4 shows that the subjective quality of the speech transcoded by the proposed algorithm is equivalent to the quality of the speech transcoded by the tandem algorithm.

4. CONCLUSION

In this paper, we proposed a novel transcoding algorithm for the AMR and EVRC speech codecs. The proposed algorithm transcodes the speech by direct parameter transformation. We also evaluated the performance of the proposed algorithm and compared it with that of the conventional

Table 4. ABX preference test results

	AMR → EVRC		EVRC → AMR	
Preference	Male	Female	Male	Female
Tandem	30%	32%	35%	27%
Proposed	40%	33%	32%	30%
No Preference	30%	35%	33%	43%

tandem algorithm. The proposed algorithm transcodes the speech by converting the parameters commonly used by both speech codecs in the parametric domain. The proposed algorithm produce equivalent speech quality to that of the tandem algorithm while requiring shorter delay and less computational complexity (50% reduction). The general idea of the proposed algorithm is not only restricted to the AMR and EVRC but is applicable to various other CELP based codecs.

5. REFERENCES

[1] 3GPP TS 26.090 V5.0.0, "AMR Speech Codec; Transcoding functions," Jun. 2002.

[2] TIA/EIA/IS-127, "Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems," 1997.

[3] Manfred R. Schroeder, Bishnu S. Atal, "Code excited linear prediction(CELP): high quality speech at very low bit rates," *Proc. of ICASSP*, pp. 937-940, 1985.

[4] R. Salami, C. Laflamme, J. P. Adoul, and D. Massaloux, "A toll quality 8 kb/s speech codec for the personal communications system (PCS)," *IEEE Trans. on Vehicular Technology*, vol. 43, no. 3, pp. 808-816, Aug. 1994.

[5] W. B. Kleijn, P. Kroon, "The RCELP Speech-Coding Algorithm," *European Trans. on Telecom.*, vol. 5, no. 5, pp. 573-582, 1994.

[6] K. K. Paliwal, B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. on Speech and Audio Processing*, vol. 1, no. 1, pp. 3–14, 1993.

[7] Sunil Lee, "Novel tandemless transcoding algorithm for AMR and EVRC speech coders," MS thesis, KAIST, 2002.

[8] 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.