An Embedded ACELP Speech Coding Based on the AMR-WB Codec

Kyung Jin Byun, Ik Soo Eo, Hee Bum Jeong, and Minsoo Hahn

ABSTRACT—This letter proposes a new embedded speech coding structure based on the Adaptive Multi-Rate Wideband (AMR-WB) standard codec. The proposed coding scheme consists of three different bitrates where the two lower bitrates are embedded into the highest one. The embedded bitstream was achieved by modifying the algebraic codebook search procedure adopted for the AMR-WB codec. The proposed method provides the advantage of scalability due to the embedded bitstream, while it inevitably requires some additional computational complexity for obtaining two different code vectors of the higher bitrate modes. Compared to the AMR-WB codec, the embedded coder shows improved speech qualities for two higher bitrate modes with a slightly increased bitrate caused by the decreased coding efficiency of the algebraic codebook.

Keywords—Speech coding, embedded speech coder, wideband speech, algebraic codebook, AMR-WB.

I. Introduction

In recent mobile communication systems, the algebraic code-excited linear predictive (ACELP) coding algorithm has been adopted for wideband speech codecs such as the Adaptive Multi-Rate Wideband (AMR-WB) standard as well as for many narrow band standard speech codecs such as G.729, the Global System for Mobile Communications (GSM) Enhanced Full Rate, Enhanced Variable Rate Coding, and the AMR codec. Up to the present, most speech codecs used in mobile communication systems have operated on a narrow bandwidth limited to 200 to 3400 Hz. However, as AMR-WB extends the audio bandwidth to 50 to 7000 Hz, it becomes possible to achieve high quality speech signals both in intelligibility and naturalness. The AMR-WB codec is the speech codec most recently standardized by the 3GPP (3rd Generation Partnership Project) for GSM and WCDMA 3G systems [1] and has been selected as the new ITU-T G.722.2 standard [2]. The AMR-WB codec is a multi-rate codec with nine different bitrates between 6.6 and 23.85 kbps. Even though it operates on nine different bitrates, the bit allocation for each codec is very similar since each codec is based on the same algorithm, i.e., the ACELP algorithm [3].

On the other hand, as multimedia communications such as ATM teleconferencing and other various Internet applications are being more widely used, packetized speech communication has become a more important issue. In this packet communication, packet losses can occur due to network congestion, exceeding delay constraints, or a buffer overflow. One of the solutions to avoid a large distortion of output speech caused by packet losses is to utilize an embedded coding scheme [4]-[8]. An embedded coding structure usually consists of a base-layer and an enhancement-layer as shown in Fig. 1.

The embedded bitstream allows the decoder to reconstruct

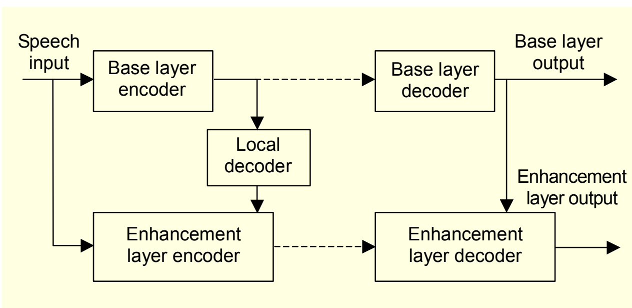

Fig. 1. Typical embedded coding structure.

Manuscript received Nov. 26, 2004; revised Mar. 07, 2005.

Kyung Jin Byun (phone: +82 42 860 5831, email: kjbyun@etri.re.kr), Ik Soo Eo (email: iseo@etri.re.kr), Hee Bum Jeong (email: hbjeong@etri.re.kr) are with Basic Research Laboratory, ETRI, Daejeon, Korea.

Minsoo Hahn (email: mshahn@icu.ac.kr) is with the School of Engineering, Information and Communications University, Daejeon, Korea.

different versions of synthetic speech at various quality levels. The base-layer bitstream guarantees the reconstruction of the original signals with a minimum acceptable quality, while the enhancement-layer bitstream enables a rather improved decoded signal quality. Therefore, an embedded coding structure usually requires an additional functional unit to implement the enhancement-layer.

In this letter, we propose an embedded ACELP speech coding structure based on the AMR-WB standard codec without any additional functional unit. The proposed embedded speech coder has three different bitrates, and the embedded bitstream can be obtained by modifying the algebraic codebook search method used in the AMR-WB codec.

II. Proposed Embedded Coding Structure

The AMR-WB is a multi-rate codec with nine different bitrates. Although the AMR-WB coder works at multiple bitrates, the bitstream for each bitrate is a fixed one. However, if the transmitted bitstream is organized in such a way that the lower bitrate mode can be embedded into the higher bitrate mode, the receiver can reconstruct speech with the available bitstream even when some parts of the highest bitrate bitstream are missed.

The AMR-WB codec operates at nine different bitrates, but the bit allocations for the seven highest bitrates, from 12.65 to 23.85 kbps, each have the same bit allocation structure; the exceptions are the algebraic codebook index and the high frequency band energy, which is only used at 23.85 kbps [1]. This implies that if the coder is able to enhance the algebraic code vector in an embedded coding manner while keeping the other parameters unchanged, an embedded coding structure can be obtained without any additional enhancement unit.

Therefore, we implemented the embedded coding structure by exploiting these similarities in the bit allocation of the AMR-WB coder. In other words, we proposed a way of evaluating the code vectors with the various resolutions by simply modifying the algebraic codebook search procedure. In view of using multi-excitation for bitrate scalability, the proposed coding scheme is similar to general multi-stage excitation coding based on an embedded coding approach [5]. However, the proposed coder can be implemented with a very simple embedded scheme by modifying the codebook search procedure adopted in the AMR-WB codec.

In the algebraic codebook of the AMR-WB coder, the subframe is divided into some tracks, and each pulse is located in these tracks for efficient modeling of the excitation signal of the subframe. For the 23.85 kbps mode, the 64 positions in each subframe are divided into four tracks where each track contains six pulses. In order to find the optimal pulse positions, the algebraic codebook is searched by maximizing the search criterion

$$Q_k = \frac{(d^t c_k)^2}{c_k^t \Phi c_k},$$

where c_k is the code vector with index k, d is the correlation vector between the target signal and the impulse response of the weighted synthesis filter, and Φ is the correlation matrix of the impulse response [1].

The AMR-WB coder at the 12.65 kbps mode, the base-layer codec of our embedded coding structure, searches the best 8-pulse code vector using a two-track-based sequential search method. While testing the possible combinations of two pulses, a limited number of potential positions of the first pulse are tested for better complexity reduction. The first two pulses are set into the positions corresponding to the maximum values in each track. The rest of the pulses are searched in pairs by searching sequentially each of the pulse pairs in the successive tracks.

Figure 2 shows the procedure to find the three optimum code vectors for the proposed embedded coding structure. In the proposed embedded coder, the algebraic codebook is searched by using the depth-first tree search method adopted in the 12.65 kbps mode AMR-WB [1]. However, while the AMR-WB coder can only get one code vector for the selected operation mode, the embedded coder can get three code vectors simultaneously by extending the search procedure of the depth-first tree search method. These three code vectors can allow the proposed speech coder to have bitrate scalability.

The detailed procedure to obtain the three code vectors in Fig. 2 can be summarized as follows. In the codebook of the lowest bitrate mode, two pulses are placed in each track giving a total

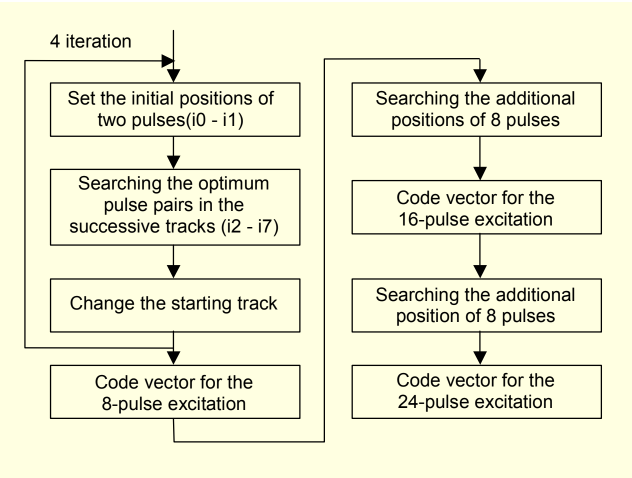

Fig. 2. Block diagram for generating three code vectors.

of eight pulses per subframe. In order to find the code vector consisting of eight pulses, two pluses are searched at the same time. Since these two pulses always correspond to consecutive tracks, the two pulses to be searched should be in tracks T0 and T1, T1 and T2, T2 and T3, or T3 and T0. Therefore, the search tree has four levels in this case. At the first level, pulses i0 and i1 are assigned to tracks T0 and T1. In these levels, no search is performed and the two pulse positions are set to the two maxima of the predefined reference signal in each track. At the second level, pulse i2 is assigned to the next consecutive track, T2, and pulse i3 is assigned to track T3. Four positions for pulse i2 are tested against all sixteen positions of pulse i3. Then, similar to the previous levels, pulse pairs corresponding to consecutive tracks are searched for. These procedures are repeated until the positions of all eight pulses are determined. The whole procedure mentioned above is repeated four times by assigning the initial two pulses to different tracks.

After all eight pulse positions are determined, we continue the search procedure in a similar manner mentioned above to find the additional eight pulse positions. Consequently, we can obtain the 16-pulse excitation having four pulses in each track. By performing the search procedure with four levels once again as shown in Fig. 2, we can finally obtain the code vector having six pulses in each track. In the additional search procedure, we did not adopt an iteration procedure for the higher bit rate mode in order to reduce the computational complexity. For the mid-rate and lowest-bit-rate modes, the embedded coding structure needs almost the same complexity as the AMR-WB codebook search at the same bit rate. For the highest bit rate mode, however, the embedded coding structure requires more computation than that of AMR-WB. As a result, it becomes possible to organize the embedded bitstream since we obtained the three kinds of code vectors.

III. Evaluation and Experimental Results

The proposed coder has the advantage of providing scalability at the cost of slightly increased computation and encoding efficiency at a higher bitrate. When encoding the pulses in a track, since some redundancy exists for the case of more than one pulse in a track, it is possible to reduce the number of bits needed for the pulse encoding [1]. However, in our embedded coder, only two pulses can be encoded at once for bitstream scalability. As a result, coding efficiency for encoding the pulses is decreased in our embedded scheme.

As shown in Table 1, in the case of two pulses per track, the required bits for encoding the pulses are the same for the AMR-WB and the proposed coders. In the case of four or six pulses per track, the embedded coder needs two and five more bits per track, respectively, due to the decreased coding efficiency.

Table 1. Required bits for pulse index encoding.

Pulses/track	AMR-WB	Embedded coding
2	9×4=36 (12.65 kbps)	9×4=36 (12.65 kbps)
4	16×4=64 (18.25 kbps)	(9+9) ×4=72 (19.85 kbps)
6	22×4=88 (23.85 kbps)	(9+9+9) ×4=108 (27.85 kbps)

The performance of the embedded coder is summarized in Table 2 in terms of the segmental signal-to-noise ratio (SNR). Although the segmental SNR may not be an adequate measurement to indicate the subjective speech quality, it still can be considered as a meaningful indicator of the relative performances between two coders. In Table 2, the performance of our coder at the lowest bitrate is equal to that of the AMR-WB coder. However, there is a slightly better quality of 0.1 and 0.2 dB in the cases of four and six pulses, respectively, because the algebraic code vectors for the higher bitrate are acquired by searching from a wider search space than that of the AMR-WB coder.

Table 2. Average segmental SNR performances.

Pulses/track	AMR-WB	Embedded coding
2	14.96 (dB)	14.96 (dB)
4	17.19 (dB)	17.26 (dB)
6	18.56 (dB)	18.76 (dB)

On the other hand, to compare the performance between the AMR-WB and the proposed coder at the same bitrate, we measured the SNR of AMR-WB at 19.85 kbps. The SNR of AMR-WB is 17.55 dB at 19.85 kbps, while that of our coder, shown in Table 2, is 17.26 dB. The better quality of AMR-WB at 19.85 kbps compared to the proposed coder reasonable since the AMR-WB coder has more pulses than the proposed coder at 19.85 kbps.

IV. Conclusion

In this letter, we proposed an embedded coding structure based on the standard AMR-WB coder. The proposed coder has three different bitrate modes, and the embedded bitstream was implemented by modifying the algebraic codebook search method used in the AMR-WB coder. Even though our coder can provide the advantage of scalability, it was implemented without any additional functional unit for evaluating two different code vectors of the higher bitrate modes. The experimental results show that our coder provides scalability while maintaining the same or better quality for all different bitrate modes with a slightly increased bitrate caused by the decreased coding efficiency of the pulse encoding in the algebraic codebook.

References

[1] 3GPP TS 26.190, *AMR Wideband Speech Codec; Transcoding Functions*, 2001.

[2] ITU-T Recommend. G.722.2, *Wideband Coding of Speech at Around 16 kbit/s Using Adaptive Multi-Rate Wideband (AMR-WB)*, 2002.

[3] B. Besette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K.Jarvinen, "The Adaptive Multirate Wideband Speech Codec (AMR-WB)," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 8, 2002, pp. 620 –636.

[4] S. Zhang and G. Lockhart, "Embedded RPE Based on Multistage Coding," *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 4, July 1997, pp. 367-371.

[5] T. Nomura, M. Iwadare, M. Serizawa, and K. Ozawa, "A Bitrate and Bandwidth Scalable CELP Coder," *Proc. ICASSP*, 1998, pp. 341-344.

[6] K. Koishida, V. Cuperman, and A. Gersho, "A 16-kbit/s Bandwidth Scalable Audio Coder Based on the G.729 Standard," *Proc. ICASSP*, 2000, pp. 1149-1152.

[7] Sean A. Ramprashad, "A Two Stage Hybrid Embedded Speech/Audio Coding Structure," *Proc. ICASSP*, 1998, pp. 337-340.

[8] R. D. De Iacovo and D. Sereno, "Embedded CELP Coding for Variable Bit-Rate between 6.4 and 9.6 kbit/s," *Proc. ICASSP*, 1991, pp. 681-684.