A BITRATE AND BANDWIDTH SCALABLE CELP CODER

Toshiyuki Nomura   Masahiro Iwadare   Masahiro Serizawa   Kazunori Ozawa

C&C Media Research Labs., NEC Corporation
4-1-1, Miyazaki, Miyamae-ku, Kawasaki, Kanagawa, 216, JAPAN

ABSTRACT

This paper proposes a flexible CELP speech coder with bitrate and bandwidth scalabilities for multimedia applications. The coder is based on multi-pulse-based CELP coding and consists of a bitrate scalable base-band coder and a bandwidth extension tool. The bitrate scalable base-band CELP coder employs multi-stage excitation coding based on an embedded-coding approach. The multi-pulse excitation codebook at each stage is adaptively produced depending on the selected excitation signal at the previous stage. The bandwidth scalability is realized by bandwidth-conversion from base-band CELP parameters to those for wideband without a widely used subband structure. The bandwidth-conversion improves base-band coding quality and expands bandwidth, simultaneously. The comparison test results show that the bitrate scalable coder is equivalent in speech quality to the fixed-bitrate CELP coder at the same bitrate for the narrowband speech. In the MOS tests, the proposed 16 kbit/s coder with the bandwidth scalability achieves equivalent coding quality to ITU-T G.722 at 56 kbit/s. The proposed coder is currently evaluated as the MPEG-4 CELP speech standard.

1. INTRODUCTION

Since the multimedia applications such as videophone and videoconference on ATM and Internet are widely used, the high-quality speech coders are highly demanded. These kinds of applications require special considerations for packet losses. One of its solution is to realize a scalable coder where the synthesized speech signal can be decoded from received packets which contain a part of the whole encoded bitstream. Furthermore, a bandwidth scalability to cover both narrowband (3.4 kHz bandwidth) and wideband (7 kHz bandwidth) speech increases the application effectiveness because wideband speech transmission provides more natural quality than a narrowband speech signal. One of standardization activities for such areas is undergoing at the ISO/IEC JTC1/SC29/WG11 (MPEG) [1].

Code excited linear predictive (CELP) coding is an efficient algorithm to realize a high-quality speech coder operating at a low bitrate. To realize a bitrate scalable coder based on CELP coding, an embedded-coding approach has been proposed[2]. The excitation signal in CELP coding is encoded by means of multi-stage coding strategy preparing a fixed-entry codebook. However, the performance is degraded at higher bitrates due to use of the fixed-entry codebook in comparison with the optimum 1-stage coding. To realize a bandwidth scalability, a subband structure has been employed [3][4][5], where both band signals are encoded by CELP-based coders. However, the audible artifacts appear because CELP coding does not keep the waveform as waveform coding and aliasing distortion due to the subband analysis is difficult to be cancelled out in the subband synthesis procedure.

This paper proposes a bitrate and bandwidth scalable speech coder based on a multi-pulse-based CELP (MP-CELP) coder [6][7]. The proposed coder simultaneously provides two kinds of scalabilities which are effective in multimedia communications. The bitrate scalability is realized retaining high coding quality by incorporating adaptive pulse-position control to produce the multi-pulse excitation codebook adaptively. A bandwidth extension tool achieves both of coding quality enhancement in base-band and bandwidth expansion. The bandwidth extension tool directly encodes the wideband speech signals without using the subband structure. The wideband signal is reproduced by decoding the whole bitstream encoded at the bandwidth extension tool and the bitrate scalable MP-CELP coder. Section 2 describes a basic structure for the MP-CELP coding. The proposed scalable CELP coder is presented in Section 3. Evaluation results are shown in Section 4. Section 5 gives the conclusion from this work.

2. MP-CELP CODING

The MP-CELP coder achieves a high coding performance by introducing a multi-pulse vector quantization [6][7] and the multi-mode coding [8]. The input signal is processed through linear predictive (LP) and pitch analysis. The LP coefficients are quantized in the line spectrum pairs (LSP) domain. The pitch delay is encoded using an adaptive codebook. The residual signal for the LP and the pitch analysis is encoded by the multi-pulse excitation scheme. The multi-pulse excitation signal is composed of several non-zero pulses. The pulse positions are restricted in the algebraic-structure codebook [9][10] and determined by an analysis-by-synthesis procedure. The pulse amplitudes are vector-quantized [6]. The gains for the pitch predictor and the multi-pulse excitation signal are normarized by the frame energy and encoded. Several modes are prepared in this gain encoding [8].

The MP-CELP coder operates at various bitrates ranging from 4 through 12 kbit/s utilizing the flexibility in multi-pulse exicitation coding [11].

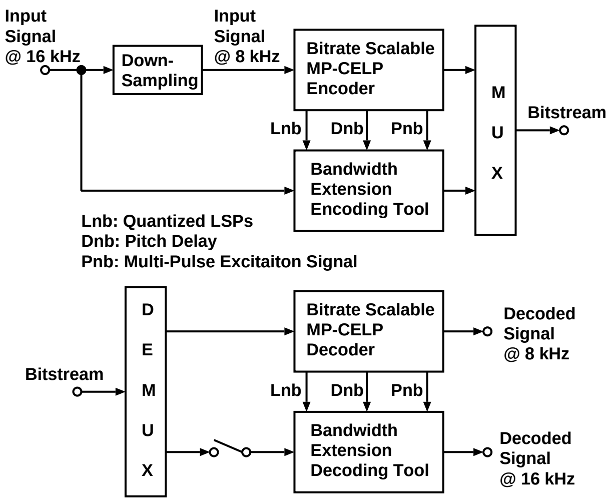

Figure 1. A block diagram of the proposed coder with both bitrate and bandwidth scalabilities.

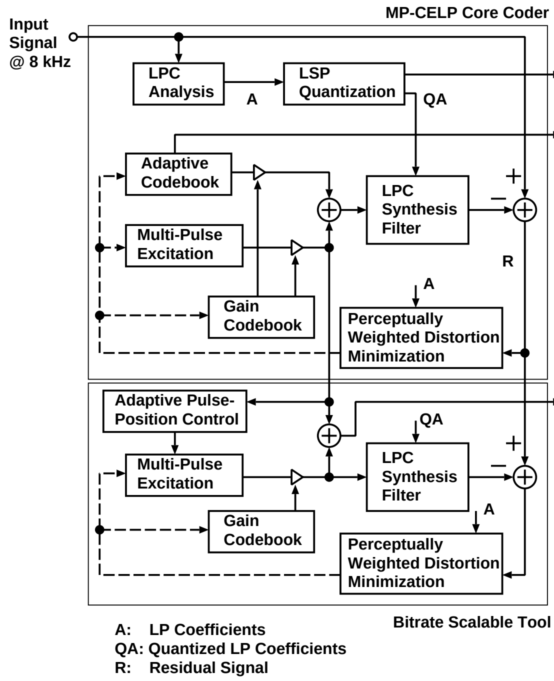

Figure 2. Bitrate scalable MP-CELP coder.

3. BITRATE AND BANDWIDTH SCALABLE CELP CODER

A block diagram of the proposed bitrate and bandwidth scalable CELP coder is shown in Figure 1. The proposed scalable coder consists of a bitrate scalable MP-CELP coder and a bandwidth extension tool. If only the bitrate scalability is necessary, the bandwidth extension tools are not required.

The bitrate scalable MP-CELP coder encodes the down-sampled 8-kHz signal. The bandwidth extension tool encodes the input signal sampled at 16 kHz utilizing the encoded parameters at the bitrate scalable MP-CELP coder. The quantized LSPs, the pitch delay and the multi-pulse excitation signal are transmitted from the bitrate scalable MP-CELP coder to the bandwidth extension tool. The bitstreams from the bitrate scalable MP-CELP coder and the bandwidth extension tool are multiplexed.

To reproduce the narrowband speech a part of bitstream is extracted and decoded by the bitrate scalable MP-CELP decoder. The wideband speech signal is obtained using the whole bitstream through both of the bitrate scalable MP-CELP decoder and the bandwidth extension decoding tool.

3.1. Bitrate Scalable MP-CELP Coder

The bitrate scalable MP-CELP coder is realized by adding a bitrate scalable tool to an MP-CELP core coder. A block diagram of the bitrate scalable MP-CELP coder is shown in Figure 2.

The bitrate scalable tool encodes the residual signal produced at the MP-CELP core coder utilizing the multi-pulse vector quantization. Adaptive pulse-position control is employed to change the algebraic-structure codebook at each excitation-coding stage depending on the encoded multi-pulse excitation at the previous stage. The algebraic-structure codebook is adaptively controlled to inhibit the same pulse-positions as those of the multi-pulse excitation in the MP-CELP core coder. The pulse positions are determined so that the perceptually weighted distortion between the residual signal and the output signal from the scalable tool is minimized. The LP synthesis and the perceptually weighted filters are commonly used for both the MP-CELP core and the scalable tool.

3.2. Bandwidth Extension Tool

The quantizers of a narrowband CELP coder and the proposed bandwidth extension tool are connected to realize both of bandwidth expansion and coding quality enhancement in base-band. A block diagram of the bandwidth extension tool is shown in Figure 3.

The bandwidth extension tool encodes the input signal at 16 kHz sampling frequency by a similar coding algorithm to the narrowband MP-CELP coder. The difference is that the LSPs, the pitch delay and the excitation signal are multi-stage encoded utilizing those in the narrowband MP-CELP coder.

3.2.1. LSP Quantization

An intraframe prediction module is incorporated in the LSP quantizer to effectively utilize the LSPs in the narrowband MP-CELP coder. A block diagram of the LSP quantizer is shown in Figure 4. The input LSPs $f_{wb}(i), i = 1, ..., N_{wb}$ are vector quantized with intraframe and interframe prediction approaches. The intraframe prediction module produces an estimated LSPs $f_{est}(i), i = 1, ..., N_{wb}$ by converting the quantized LSPs $\hat{f}_{nb}(i), i = 1, ..., N_{nb}$ obtained in the narrowband MP-CELP coder. Furthermore, an interframe moving average predictive VQ [12] is also employed for more accurate prediction. The quantized LSPs in the band extension tool are given by:

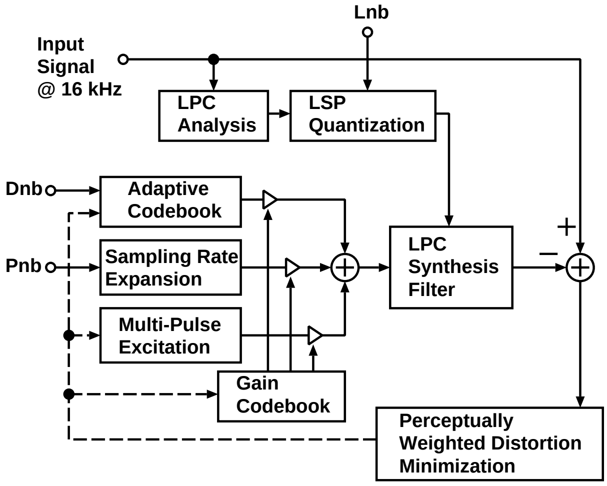

Figure 3. Bandwidth extension tool.

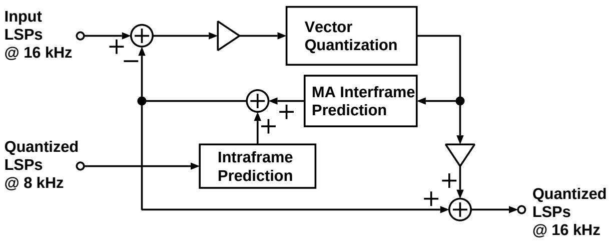

Figure 4. A block diagram of LSP quantizer.

$$\hat{f}_{wb}(i) = \sum_{p=0}^{P} a_p(i)c_p(i) + f_{est}(i), \quad i = 1, ..., N_{wb},$$

$$f_{est}(i) = \begin{cases} b(i)\hat{f}_{nb}(i), & \text{for } i = 1, ..., N_{nb}, \\ 0.0, & \text{for } i = N_{nb}+1, ..., N_{wb} \end{cases}$$

where $a_p(i)$ are interframe prediction coefficients and P is the order of interframe prediction. $c_p(i)$ are the quantized values of the prediction error vector at the p-th previous frame. Estimation coefficients $b(i)$ are used in intraframe prediction, where the narrowband quantized LSPs are transformed to those in the wideband frequency domain. This scheme quantizes not only higher-band LSPs but also the LSPs in the narrowband more acculately.

3.2.2. Pitch Delay and Multi-Pulse Excitation Signal Encoding

Two pathes from the narrowband MP-CELP coder are prepared to utilize the pitch delay and the multi-pulse excitation signal. The excitation signal is represented by a linear combination of the adaptive codevector and two kinds of multi-pulse excitation signals scaled by their respective gains. The adaptive codevector is produced from a block of the past excitation signal with the pitch delay. The pitch delay is decided in the range around an estimated pitch delay which is same as the pitch delay selected in the narrowband MP-CELP coder. One of the two multi-pulse excitation signals (MP1) is obtained by sampling-rate expansion of the multi-pulse excitation signal used in the narrowband MP-CELP

Table 1. Bit allocation for the proposed coder (bits/frame).

	MP-CELP Core Coder	Bitrate Scalable Tool	Band Extension Tool
Mode	2	–	–
Frame Energy	6	–	–
LSP	22	–	32
Pitch Delay	8 × 2	–	3 × 4
Multi-Pulse	31 × 2	16 × 2	28 × 4
Gain	6 × 2	4 × 2	(7+4) × 4
Bitrate [kbit/s]	6.0	2.0	10.0

coder. The another multi-pulse excitation signal (MP2) consists of several non-zero pulses. The pulse positions and amplitudes are determined so as to minimize the perceptually weighted distortion by a similar method to that of the narrowband MP-CELP coder. The gains for the adaptive codevector and the MP2 excitation signal are vector quantized. The gain for the MP1 excitation signal is scalar quantized. These quantization operations are achieved by minimizing the perceptually weighted distortion.

4. PERFORMANCE EVALUATION OF SCALABLE CELP CODER

The performance of the proposed scalable coder was evaluated subjectively and objectively. The detailed bit allocations for the bitrate and bandwidth scalable coder are shown in Table 1. The MP-CELP core coder and the two tools operate at 20-ms frame. Each frame consists of two subframes for the MP-CELP core and the bitrate scalable tool. The band extension tool has four subframes in 20-ms frame.

4.1. Bitrate Scalable Coder

The effectiveness of the adaptive pulse-position control was evaluated using average segmental SNRs. Paired comparison tests were also conducted. The evaluated 12-kbit/s bitrate scalable coder was realized by adding three stages of the 2-kbit/s bitrate scalable tool to the 6-kbit/s MP-CELP core coder. The narrowband speech signals at 8 kHz sampled were used.

The average segmental SNRs for the bitrate scalable coders with and without the adaptive pulse-position control are shown in Table 2. The results indicate that the segmental SNRs for the coder with the adaptive pulse-position control are 0.10, 0.26 and 0.54 dB higher than those without the adaptive control at 8, 10 and 12 kbit/s, respectively.

In the paired comparison tests, the quality of the proposed coder at 8, 10 and 12 kbit/s were evaluated against those of the fixed-bitrate MP-CELP coder at 8.3, 10.3 and 12 kbit/s, respectively. The 5-point relative score was used to grade the relative quality. The comparison test results are summarized in Table 3. The results show that the coding quality of the proposed coder is equivalent to those of the fixed-bitare MP-CELP coder at respective bitrates.

4.2. Bandwidth Scalable Coder

Subjective performance for the bandwidth scalable MP-CELP coder was evaluated by a mean opinion score (MOS) test. The

Table 2. Average segmental SNR performance for the bitrate scalable coders.

Bitrate (kbit/s)	With APPC	Without APPC
6+2	11.78 (dB)	11.68 (dB)
6+2+2	12.49 (dB)	12.23 (dB)
6+2+2+2	13.11 (dB)	12.57 (dB)

APPC: Adaptive Pulse-Position Control

Table 3. Paired comparison test results for the narrowband speech.

Tested Coder vs Reference Coder	Averaged Score	95 % Confidence Interval
Scalable 8 kbit/s vs Fixed 8.3 kbit/s	-0.10	0.09
Scalable 10 kbit/s vs Fixed 10.3 kbit/s	-0.12	0.08
Scalable 12 kbit/s vs Fixed 12 kbit/s	-0.02	0.08

Table 4. Equivalent Q values for the wideband speech.

Coder Type	Equivalent Q values (dB)
Bandwidth Scalable MP-CELP (16 kbit/s)	38.5
Subband MP-CELP (16 kbit/s)	37.1
Fullband MP-CELP (16 kbit/s)	40.0
G.722 (48 kbit/s)	34.1
G.722 (56 kbit/s)	39.1
G.722 (64 kbit/s)	39.5

evaluated bandwidth scalable coder is composed of the 6-kbit/s MP-CELP core coder and the 10-kbit/s bandwidth extension tool. Fullband [13] and subband coders based on MP-CELP were included, where the coders operate at 16 kbit/s with 10-ms frame. ITU-T Rec. G.722 at 48, 56 and 64 kbit/s were also evaluated. The wideband speech signals sampled at 16 kHz were used as test materials. Table 4 shows the test results by equivalent Q values based on MOS for the modulated noise reference unit (MNRU). The Q value of the proposed coder does not reach that of the fullband MP-CELP coder due to the penalty for the bandwidth scalability. But the coder achieves better coding quality than the subband MP-CELP coder. The results also indicate that the proposed bandwidth scalable coder achieves equivalent coding quality to that of G.722 at 56 kbit/s.

5. CONCLUSION

A bitrate and bandwidth scalable CELP coder has been proposed. This coder consists of a bitrate scalable MP-CELP coder and a bandwidth extension tool. The bitrate scalability is achieved with adaptive pulse-position control which reduces quality degradation caused by the multi-stage excitation coding. The bandwidth scalability is realized by bandwidth extension tools for CELP parameters without the subband structure. The bandwidth extension tool achieves coding quality enhancement in base-band as well as bandwidth expansion. The comparison test results show that the coding quality of the bitrate scalable coder is equivalent to those of the fixed-bitare CELP coders at the same bitrates. The MOS test results indicate that the coding quality of the 16 kbit/s bandwidth scalable coder is equivalent to that of ITU-T G.722 at 56 kbit/s.

The proposed coder is suitable for ATM and Internet communications when combined with quality of service (QoS) control. Therefore, MPEG is currently evaluating this coder as the MPEG-4 CELP speech standard. Moreover, the proposed schemes are applicable to other CELP coders. For example, the bandwidth expansion tool is easily merged with ITU-T G.729 and GSM EFR. This integeration will produce new communication services which change the speech quality and the bitrate depending on the transmission condition.

ACKNOWLEDGEMENTS

The authors would like to thank Akihiko Sugiyama and Atsushi Murashima for their valuable suggestions.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11/N1745, "Working Draft of ISO/IEC 14496-3 MPEG-4 Audio Ver. 3.0," 1997.

[2] R. Drogo de Iacovo, et al., "Embedded CELP Coding for Variable Bit-Rate between 6.4 and 9.6 kbit/s," Proc. ICASSP'91, pp.681-684, 1991.

[3] CCITT Rec. G.722, "7 kHz Audio Coding within 64 kbit/s," CCITT Red Book, 1984.

[4] R. Drogo de Jacovo, et al., "Some Experiments of 7 kHz Audio Coding at 16 kbit/s," Proc. ICASSP'89, pp.192-195, 1989.

[5] A. Kataoka, et al., "A 16-kbit/s Wideband Speech Codec Scalable with G.729," Proc. Eurospeech'97, pp.1491-1494, 1997.

[6] S. Taumi, et al., "Low-delay CELP with multi-pulse VQ and fast search for GSM EFR," Proc. ICASSP'96, pp.562-565, 1996.

[7] K. Ozawa, et al., "MP-CELP Speech Coding Based on Multi-Pulse Vector Quantization and Fast Search," IEICE, Vol.J79-A, No.10, pp.1655-1663, 1996 (In Japanese).

[8] K. Ozawa, et al., "M-LCELP Speech Coding at 4 kbit/s with Multi - mode and Multi - codebook," IEICE Trans. Commun., Vol.E77-B, No.9, pp.1114-1121, 1994.

[9] C. Laflamme, et al., "16 kbps Wideband Speech Coding Technique Based on Algebraic CELP," Proc. ICASSP'91, pp.13-16, 1991.

[10] ITU-T Rec. G.729, "Coding of speech at 8-kbit/s using conjugate structure algebraic code-excited linear prediction (CS-ACELP)," COM 15-152-E, 1995.

[11] T. Nomura, et al., "Proposal of compression algorithm with rate control for MPEG-4/Audio core experiments," ISO/IEC JTC1/SC29/WG11/M1509, 1996.

[12] H. Ohmuro, et al., "Coding of LSP Parameters Using Interframe Moving Average Prediction and Multi-Stage Vector Quantization," Proc. IEEE Speech Coding Workshop, pp.63-64, 1993.

[13] M. Serizawa, et al., "A 16 kbit/s Wideband CELP Coder with a High-Order Backward Predictor and Its Fast Coefficient Calculation," Proc. IEEE Speech Coding Workshop, pp.107-108, 1997.