AN ADAPTIVE MULTI-RATE SPEECH CODEC BASED ON MP-CELP CODING ALGORITHM FOR ETSI AMR STANDARD

Hironori Ito, Masahiro Serizawa, Kazunori Ozawa and Toshiyuki Nomura

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

ABSTRACT

This paper proposes a speech codec based on the Multi-Pulse based CELP (MP-CELP) coding and a convolutional coding algorithms for the ETSI Adaptive Multi-Rate (AMR) standard. The codec operates at several speech coding rates, maintaining a fixed gross rate including speech and channel coding for the Full-Rate (FR) and Half-Rate (HR) channel modes. MP-CELP has great features of easily changing the speech coding rate by controlling the parameters such as the number of pulses and other parameters. Subjective tests show that the proposed AMR codec in the FR channel mode achieves higher performance than that of the Enhanced FR codec, and the proposed codec in the HR channel mode gives a comparable coding quality to that by the Full-Rate codec, by selecting an optimal coding rate for each channel condition. T-tests based on the test results also show that the proposed speech codec meets about 80 % of the seventeen requirements, which are selected from the AMR standard study report. Therefore, the proposed codec is promising for the AMR standard.

1. INTRODUCTION

Recently, several standards for speech codecs have been established for the GSM (Global System for Mobile Communication) cellular system by ETSI (the European Telecommunications Standards Institute). The codecs, which consist of a speech and a channel codecs at fixed coding rates, operate in two different channel modes: the Full-Rate at 22.8 kbit/s and the Half-Rate at 11.4 kbit/s. The FR codec, based on the RPE-LTP speech coding algorithm [1], has been used since the GSM digital cellular system started. The standard for the HR codec, based on the EVSELP speech coding algorithm [2], was then established to cope with an increasing number of cellular-telephone subscribers. Another codec in the FR channel mode, an enhanced FR (EFR) codec[3] based on the ACELP speech coding algorithm , has also been standardized later to meet the needs by subscribers for higher speech quality.

Currently, ETSI is working on a standard for an Adaptive Multi-Rate (AMR) codec. The AMR standardization has two goals [4]. The first one is significantly improved quality in the HR channel mode which enables twice as much the capacity in the FR channel mode. The other is increased robustness to high channel-error rates in the FR channel mode. These goals require a high performance codec, which controls the speech and the channel coding rates according to the channel-error condition and the traffic loading.

An 11 kbit/s codec based on the Multi-Pulse based CELP (MP-CELP) algorithm [5],[6] achieves equivalent performance to that of the 12.2 kbit/s EFR codec for clean speech. Its performance is enhanced thanks to a novel multi-pulse vector-quantization (MP VQ) [5] with combination search. Furthermore, MP-CELP algorithm can easily realize a multi-rate codec by changing the number of pulses in the MP excitation [7]. Therefore, an AMR multi-rate codec based on the MP-CELP algorithm is suitable as the basis for the AMR standard codec.

This paper proposes an adaptive multi-rate MP-CELP speech codec for the AMR standard. It is based on the multi-rate MP-CELP coding and a convolutional channel coding algorithms, where the gross coding rate is fixed in each channel mode, FR or HR. In the following section, a new adaptive multi-rate speech codec is proposed. In Section 3, subjective test results are shown and discussed on the AMR standard.

2. PROPOSED AMR MP-CELP CODEC

The proposed AMR codec integrates a multi-rate speech coding and a convolutional channel coding algorithms. It has a capability of adaptively controlling of each coding rate according to the channel condition. The coding rate control can be carried out based on metrics, e.g. Receive Quality Estimates (RXQUAL) and Receive Carrier Estimates (RXLEV), which are under discussion at the ETSI AMR meetings. Therefore, this paper does not discuss adaptive rate control based on the metrics and concentrates on speech and channel coding algorithms.

2.1. Multi-Rate MP-CELP Algorithm

The proposed multi-rate speech coding algorithm operates at multiple bitrates with a single algorithm by switching the bit allocation and the subframe length. A blockdiagram of the Multi-Rate MP-CELP coding is shown in Figure 1.

The MP-CELP coding algorithm provides high performance at any coding rate. Its coding performance is mainly enhanced by the multi-pulse VQ, where the excitation signal is modeled by multiple pulses. The pulse positions are coded based on the

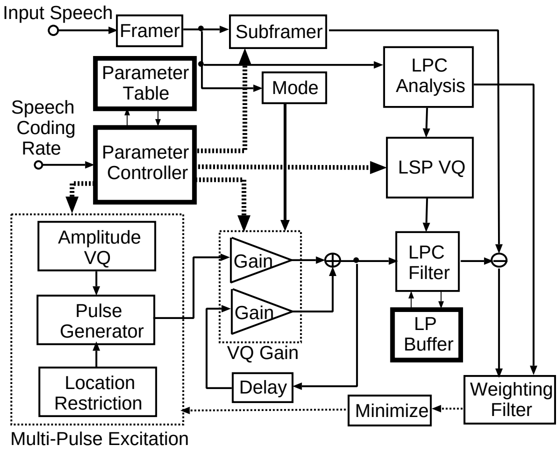

Figure 1: Blockdiagram of Multi-Rate MP-CELP Coding.

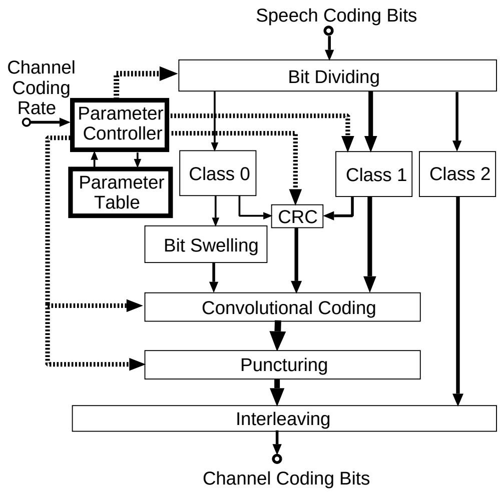

Figure 2: Blockdiagram of Channel Coding.

algebraic codebook structure [8] to reduce the coding rate. It reduces the complexity by adopting the tree coding [9]. The pulse amplitudes are vector-quantized. In order to improve the performance, a combination search between the multiple sets of pulse positions and the pulse amplitude codevectors is employed.

To realize the multi-rate codec, the bit allocation and the subframe length for each coding rate are stored in a table and used according to the specified speech coding rate. The same buffer for the LP synthetic filter is also used at all the coding rates to avoid any annoying noise caused by a change of the buffer.

To further improve the coding performance, the gains for the pitch lag and the multi-pulse excitation are coded by the multi-mode method [10] after gain normalization by the corresponding frame energy. LP coefficients are transferred to LSPs (Line Spectral Pairs) and are vector-quantized based on Moving-Average (MA) predictive VQ [11]. The pitch lag is differentially coded [12].

Table 1: Coding rates for speech and channel coding with bit allocation for speech coding parameters.

Channel mode		FR		HR	
Gross Coding Rate[kbit/s]		22.8		11.4	
Speech Coding Rate [kbit/s]		10.6	8		5.5
Channel Coding Rate [kbit/s]		12.2	14.8	3.4	5.9
Speech Coding Bit Alloc. [bit]	Mode	2			
	LSP	22			16
	Pitch Lag	26			13
	Excitation	128	80		62
	Gain VQ Frame Energy	28	24		12

2.2. Channel Coding Algorithm

A channel coding algorithm with the bit swelling technique [13] is used at all coding rates. A blockdiagram of the channel coding is shown in Figure 2. The bit allocation for channel coding is stored in a table and used according to the specified channel coding rate. Speech coding bits are first divided into three classes [14]; Classes 0, 1 and 2. Then, three bits of Class 0, which has most sensitive to bit errors, are swollen into six-bit representation for stronger errors protection. Cyclic Redundancy Check (CRC) codes are applied to the several most sensitive bits in Classes 0 and 1 for error detection. Finally, the Class 0 bits after bit swelling, the Class 1 bits and the CRC bits are coded with a convolutional code followed by puncturing [15] and interleaving [16] operations. The Class 2 bits are not protected. If the error is detected by the CRC codes, the missing data are extrapolated to recover the current speech frame from the decoded past frames.

3. PERFORMANCE EVALUATION

The proposed AMR codec operates at gross rates of 22.8 and 11.4 kbit/s in the FR and the HR channel modes, respectively. The speech and the channel coding rates used in speech and channel coding are shown in Table 1. The bit allocation for the speech coding parameters is also included. With the frame length of 20 msec, a 5-msec subframe is used at speech coding rates of 10.6 and 8 kbit/s and a 10-msec subframe at a speech coding rate of 5.5 kbit/s. There is no look-ahead delay due to LPC analysis, resulting in a 20-msec coding algorithmic delay.

3.1. Error Correction Capability

The error correction capability was evaluated using 6400-frame speech coding bits under three channel-error conditions: C/I (Carrier to Interfere ratio) = 10 dB (EP1), C/I = 7 dB (EP2) and C/I = 4 dB (EP3). Error rates measured for Classes 0 and 1 are shown in Table 2. The error rates of Classes 0 and 1 after error correction were lower than 1 %, except for some conditions under EP2 and EP3. In such cases, the error rates were reduced to less than 1 %

Table 2: Error rates after error correction (EC) [%]

Error Pattern		EP1	EP2	EP3
Error rate before EC		4.48	7.55	12.1
FR channel mode	10.6 kbit/s	0.00	0.13	2.27
	8 kbit/s	0.00	0.01	0.87
Error rate before EC		4.51	7.60	12.1
HR channel mode	8 kbit/s	0.10	1.23	6.02
	5.5 kbit/s	0.05	0.33	2.37

and 3 % in the FR and the HR channel modes, respectively, since the proposed AMR codec selects a higher channel coding rate to cope with the errors.

3.2. Subjective Speech Quality

3.2.1. Test Conditions

Subjective tests for speech with and without background-noise were conducted according to the AMR qualification test plans [17],[18] discussed at the AMR meetings. The tests were simplified with respect to the numbers of speech samples and channel-error patterns. Sixteen Japanese speech samples uttered by two male and two female talkers were used. The duration of each sample was 8 seconds. All tests were carried out with 24 listeners using telephone handsets. Modulated Noise Reference Unit (MNRU) signals at 5, 10, 15, 25, 30 and 35 dB, and the original speech were included.

The performance for clean speech were tested by ACR (Absolute Category Rating) with a 5-point scale under error-free (EP0), EP1, EP2 and EP3 in the FR and the HR channel modes. The EFR and the FR codecs under EP0, EP1, EP2 and EP3, and ITU-T G.728 codecs under EP0 were included as reference codecs. The reference codecs without channel coding operate at the speech coding rates of 12.2, 12.2 and 16 kbit/s, respectively.

The performance for speech with background-noise was also tested with street noise with S/N (Signal to Noise ratio) = 15 dB and vehicle noise with S/N = 10 and 15 dB. DCR (Degradation Category Rating) with a 5-point degradation scale was employed. EP0 and EP1 were used for the proposed codec in the FR channel mode, and EP0 in the HR channel mode. The EFR codec under EP0 and the FR codec under EP0 and EP1, and ITU-T G.729 codec at EP0 were included as reference codecs. 8 kbit/s was the speech coding rate for G.729 codec without channel coding.

3.2.2. Test Results

Subjective evaluation results in the FR and the HR channel modes are shown in Figure 3 (a) and (b), respectively. In the FR channel mode, the proposed AMR codec at a fixed speech coding rate of 10.6 kbit/s achieved comparable or superior performance to that of the EFR codec, which operated at a speech coding rate of 12.2 kbit/s. At a speech coding rate of 8 kbit/s, the proposed AMR codec gave higher coding quality than that by the codec at 10.6 kbit/s under EP2 and EP3. The performance under EP1 and EP2 were comparable to that of G.728 codec. Thus, the proposed codec achieves higher coding quality than those by the reference codecs at a fixed single coding rate by selecting an optimal coding rate for each channel-error condition. In the HR channel mode, the performance of the proposed codec was equal to or higher than that of the FR codec by selecting an optimal coding rate for each channel-error condition. The proposed codec at a speech coding rate of 8 kbit/s also gave comparable performance to that of G.728 codec under EP0.

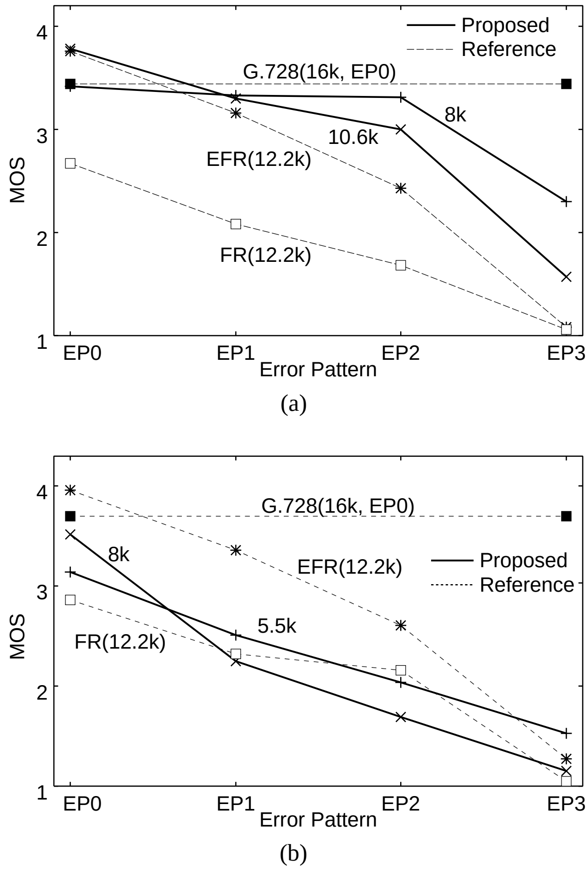

Figure 3: Subjective results in (a) the FR channel mode and (b) the HR channel mode.

The results for speech with background-noise are shown in Figure 4. In the FR channel mode, the proposed codec at a speech coding rates of 10.6 and 8 kbit/s achieved comparable performance to those of the EFR and the FR codecs except for the performance for the vehicle noise under EP0. In the HR channel mode, the coding quality by the proposed codec at a speech coding rate of 8 kbit/s was higher than that by G.729. But the quality at a speech coding rate of 5.5 kbit/s is lower than that by G.729.

T-tests based on these results also show that the proposed speech codec meets about 80 % of the seventeen requirements, which are selected from the AMR standard study report. The codec did not meet the requirements under a high bit-error condition in

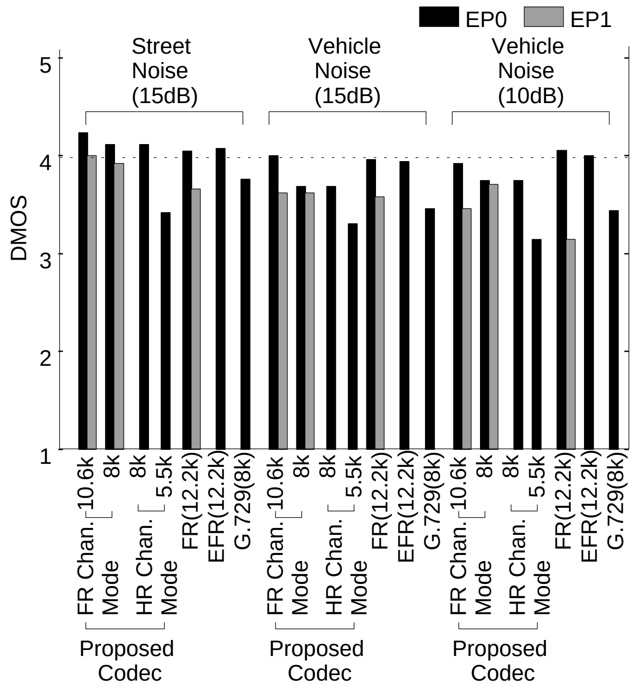

Figure 4: Subjective results for noisy speech.

the FR channel mode and a background-noise condition in the HR channel mode.

4. CONCLUSION

A speech codec based on the Multi-Pulse based CELP coding and a convolutional coding algorithms has been proposed for the ETSI AMR standard. Subjective tests show that the proposed AMR codec operating at a given optimal coding rate achieves higher performance than those of the EFR codec in the FR channel mode. It also provides a comparable performance to that of the FR codec in the HR channel mode. However, further studies are needed to improve the coding performance under a high bit-error condition in the FR channel mode and a background-noise condition in the HR channel mode from a view point of the AMR requirements.

Acknowledgment

The authors would like to thank Dr. Akihiko Sugiyama, Masao Ikekawa and Hiromi Kumagai for valuable suggestions.

REFERENCES

[1] J. Natvig, "Pan-European speech coding standard for digital mobile radio," Speech Communication, pp.113-123, 1988.

[2] ETSI ETS 300 581-2, "European digital cellular telecommunications System; Half rate speech Part 2: Half rate speech transcoding", Nov. 1995.

[3] ETSI TC-SMG, "Digital cellular telecommunications system; Performance characterization of the GSM Enhanced Full Rate(EFR) speech codec(GSM 06.55)," Aug. 1996.

[4] ETSI TC SMG, "Adaptive multi-rate(AMR) study phase report, version 1.0," no.23, Oct. 1997.

[5] K. Ozawa, et al., "MP-CELP speech coding based on multipulse vector quantization and fast search," IEICE Trans. Vol. J79-A, No.10, pp.1655-1663, Oct. 1996 (in Japanese).

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

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

[8] R. Salami, et al., "Description of the proposed ITU-T 8 kb/s speech coding standard, " IEEE Workshop on Speech Coding for Telecommunications, pp.3-4, Sep. 1995.

[9] T. Nomura, et al., "Efficient pulse search methods in CELP," Proc. ASJ, pp. 311-312, March 1996 (in Japanese).

[10] K. Ozawa, et al., "M-LCELP Speech Coding at 4 kb/s with Multi-Mode and Multi-Codebook ," IEICE Trans. On Communications, Vol.E77-B, No.9, pp.1114-1121, Sep. 1994.

[11] T. Nomura, et al., "Vector quantization of LSP parameters using adaptive prediction," National Meeting of ASJ, 1-5-7, pp.245-246, Oct. 1994(in Japanese).

[12] M. Serizawa, et al., "M-LCELP speech coding at 4kbps," Proc. IEEE Workshop on Speech Coding for Telecommunications, pp.47-48, Sep. 1993.

[13] M. Ikekawa, et al., "Effective channel coding combined with low bit-rate speech coding for digital cellular system," Proc. IEEE VTC, Vol. 3, pp1685-1689, 1994.

[14] C.-E. W. Sundberg and N. Seshadri, "Digital cellular systems for North America," Proc. IEEE GLOBECOM'90, 533-537, 1990.

[15] Cain, J.B et al.,"Punctured convolution codes of rate (n-1)/n and simplified maximum likelihood decoding, IEEE Trans. Inform. Theory, Vol.IT-25, pp.97-100 1979.

[16] G. C. Clark, Jr. et al.,"Error-Correction Coding for Digital Communications,"Plenum Press. Sec. 8.3 1981.

[17] ETSI SMG11, "AMR Qualification Test Plan - Experiments 1a & 1b. The effect of errors under clean speech conditions," BT Laboratories, TD 94/97 Aug. 1997.

[18] ETSI SMG11, "AMR Qualification Test Plan - Experiments 2a & 2b. The effect of background noise," TD 95/97 Aug. 1997.