HIGH QUALITY MUTI-PULSE BASED CELP SPEECH CODING AT 6.4 KB/S AND ITS SUBJECTIVE EVALUATION

Kazunori Ozawa and Masahiro Serizawa

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

ABSTRACT

This paper proposes an MP-CELP (Multi-Pulse-based CELP) speech coding at 6.4 kb/s with 10 ms frame. In MP-CELP, amplitudes or signs of multi-pulse excitation are simultaneously vector quantized (VQ). A combination search between multiple pulse location candidates and VQ codebook remarkably improves the quantization performance. In order to improve speech quality for background noise conditions, an adaptive pulse location restriction method is developed. The subjective evaluation results show that speech quality for 6.4 kb/s MP-CELP is higher than that for G.726 at 32 kb/s and is equivalent to that for 6.3 kb/s G.723.1 with 30 ms frame in clean speech and tandem conditions. For background noise conditions, the adaptive pulse location restriction significantly improves MOS value by 0.9. The speech quality is equivalent to that for G.723.1, but still does not reach to that of 24 kb/s G.726, except interference talker condition.

1. INTRODUCTION

In the near future, mobile multi-media communications services is expected to be provided via IMT (International Mobile Telecommunication)-2000. In that services, high quality and short delay speech communication which is as good as wire-line quality and flexibility which can provide several kinds of source coding bit rates in accordance with transmission channel capacity and bit error rates are very important issues.

For high quality speech communication, ETSI standardized Enhanced Full-Rate codec [1] whose source coding bit rate was 12.2 kb/s. ITU-T has recently standardized 8 kb/s CS-ACELP (Conjugate Structure Algebraic CELP) [2] as G.729 speech codec. It can provide high-quality speech at 8 kb/s. In mobile multi-media applications, needs for lower bit rate speech coding with short delay are increasing, while keeping high-speech quality. To provide one of solutions for such demands, ITU-T has started to study lower bit rate extension of G.729 with 10 ms frame and determined the speech quality requirements [3]. However, at bit rates lower than 8 kb/s, it is difficult to meet the requirements with short delay especially for background noise conditions.

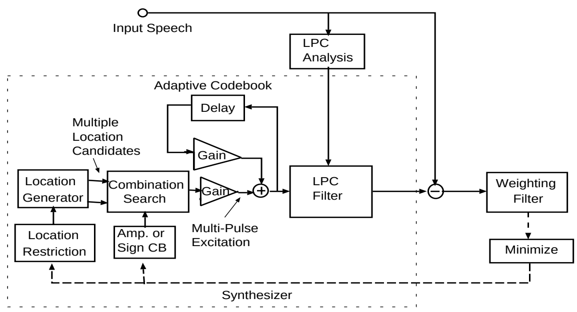

Figure 1: MP-CELP encoder operation principle.

The authors have proposed MP-CELP (Multi-Pulse-based CELP) speech coding [4], [5]. In MP-CELP, amplitudes or signs for multi-pulse excitation are simultaneously vector quantized and combination search between multiple pulse location candidates and VQ codebook enhances performance. Speech quality for 11 kb/s MP-CELP with 10 ms frame is higher than that for G.728 (16 kb/s LD-CELP) [6].

This paper proposes 6.4 kb/s speech coding with 10 ms frame, based on MP-CELP technique. In order to improve speech quality for background noise conditions, the adaptive pulse location restriction method is developed. The subjective quality evaluation results under several conditions are presented.

2. MULTI-PULSE BASED CELP

The operation principle for MP-CELP encoder is shown in Fig. 1. In this method, excitation signal is represented by multi-pulse [7], [8]. The amplitudes or signs of all multi-pulse in a sub-frame are simultaneously vector quantized to icassp98 improve the quantization performance [4], [5]. The pulse locations are restricted based on the algebraic structure which is the same as that in G.729 [2] to reduce both transmission bit rate and location search complexity.

2.1. Multi-Pulse VQ Excitation

In MP-CELP, the excitation signal $v(n)$ is represented as:

$$v(n) = \gamma \sum_{k=1}^{M} g_{kj} \delta(n - m_k), \quad (1)$$

where $\delta(0) = 1$ and $\delta(i) = 0$ for $i \neq 0$. M is the number of pulses in a sub-frame. g_{kj} and m_k are the kth amplitude of the jth amplitude codevector and the location of the kth pulse, respectively. γ shows a gain.

2.2. Improvement by Amplitude VQ

Pulse locations are searched for using fast search method based on tree coding [9] with the pre-determined pulse amplitudes. Then, multiple location candidates are selected. To improve the amplitude VQ performance, a combination search between multiple pulse location candidates and amplitude codebook (CB) is carried out. The optimal combination between the pulse location candidate and the amplitude codevector is selected so as to minimize the following distortion,

$$D_j = \sum_{n=0}^{N-1} \left[x_w(n) - \gamma \sum_{k=1}^{M} g_{kj} h_w(n - m_k) \right]^2, \quad (2)$$

where N is the number of samples in a sub-frame and $x_w(n)$ is the perceptually weighted speech signal. $h_w(n)$ shows the impulse response of synthesis filter with the perceptual weighting.

Minimization of D_j in Eq. (2) can be replaced to maximization of the value D'_j:

$$D'_j = C_j^2 / E, \quad (3)$$

where

$$C_j = \sum_{k=1}^{M} g_{kj} d(m_k), \quad (4)$$

$$E_j = \sum_{k=1}^{M} g_{kj}^2 \phi(m_k, m_k) + 2 \sum_{k=1}^{M} \sum_{i=k+1}^{M} g_{kj} g_{ki} \phi(m_k, m_i), \quad (5)$$

$$d(n) = \sum_{i=n}^{N-1} x_w(i) h_w(i - n) \quad (6)$$
$$n = 0, ..., N-1,$$

Table 1: SNR_{seg} v.s. the number of pulse location candidates N_p

N_p	$SNR_{seg}(dB)$		
	Amplitude CB		Sign CB
	7 bits	8bits	7bits
1	17.45	17.55	17.60
2	17.78	17.88	17.81
4	17.97	18.08	17.91
8	18.13	18.25	18.02
16	18.26	18.33	18.09

Table 2: An example of possible pulse locations

Pulse	Location
m_1	0,5,10,15,20,25,30,35
m_2	1,6,11,16,21,26,31,36
	2,7,12,17,22,27,32,37
m_3	3,8,13,18,23,28,33,38
	4,9,14,19,24,29,34,39

$$\phi(p, q) = \sum_{n=\max(p,q)}^{N-1} h_w(n-p) h_w(n-q), \quad (7)$$
$$p, q = 0, ..., N-1.$$

Table 1 shows the SNR_{seg} performance against the number of the pulse location candidates N_p, where sub-frame length is 5 ms and the number of pulse M is 7. It is shown that the combination search improves SNR_{seg} by 0.8 dB compared to the method without the combination search (N_p=1), when the amplitude codebook is used. In the case of sign CB, SNR_{seg} improvement of 0.5 dB is achieved when the number of candidates is 16.

By using the sign CB, the excitation structure is similar to that of G.729 [2], because G.729 excitation is considered to be without the combination search.

3. 6.4 KB/S MP-CELP STRUCTURE

In order to keep short delay, 10 ms frame size and 5 ms sub-frame size should be maintained. In this case, the number of pulses per sub-frame should be reduced to 3, and only 1 bit sign codebook (3 dimensional) should be used to quantize signs for all of three pulses, when the simulation conditions in Table 5 is used. An example of possible pulse locations is shown in Table 2.

Table 3: Four kinds of 1 bit sign codebook patterns

Index	1 bit Sign Codebook Pattern
(a)	(+,+,+),(-,-,-)
(b)	(+,-,+),(-,+,-)
(c)	(+,+,-),(-,-,+)
(d)	(+,-,-),(-,+,+)

Table 4: SNR_{seg} performance comparison

Index	Condition	$SNR_{seg}(dB)$
(a)	N_p = 16	11.77
(b)	N_p =16	12.26
(c)	N_p =16	13.00
(d)	N_p =16	12.64
(e)	(c) + N_p =1	9.82
(g)	3 bits Sign CB	14.81

3.1. 1 BIT SIGN CODEBOOK PERFORMANCE

Performances of four knids of 1 bit sign codebook patterns, given in Table 3, are compared. Table 4 shows SNR_{seg} performances. In the Table, performance using 3 bits sign codebook is also presented as a reference. In this case, bit rate is increased by 0.4 kb/s.

From Table 4, index (c) provides the best SNR_{seg} performance among 4 kinds of patterns compared. Thanks to the combination search, SNR_{seg} is improved by about 3 dB when the candidate number is 16 (Table 4 (c)), compared to no combination search (Table 4 (e)).

3.2. ADAPTIVE PULSE LOCATION RESTRICTION

At 6.4 kb/s, both the number of multi-pulse and the size of the sign codebook are small as described in the previous sections, and not sufficient to represent noise-like signal such as various kinds of background noise which has relatively low correlations. In order to improve the performance for noise-like signals, an adaptive pulse location restriction is developed. Figure 2 shows a blockdiagram for MP-CELP excitation part including the adaptive pulse location restriction.

In the adaptive pulse location restriction, possible pulse locations for voiced speech and those for unvoiced speech are different. For voiced speech, pulse location restriction shown in Table 2 is employed. For unvoiced speech, different and more strong pulse location restriction is used to represent noise-like signals. In order to avoid increase of V/UV decision information, V/UV decision is carried out by using the quantized adaptive codebook gain in the previous

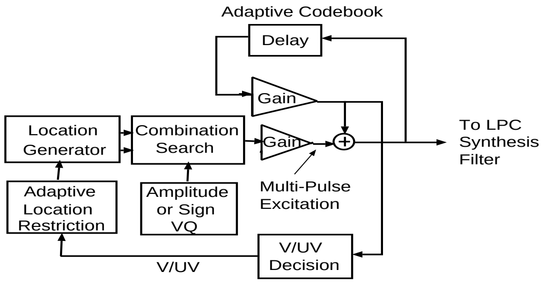

Figure 2: Blockdiagram for MP-CELP excitation part including adaptive pulse location restriction.

Table 5: 6.4 kb/s MP-CELP simulation conditions

Parameter	Number
Frame (ms)	10
Sub-Frame (ms)	5
LSP VQ (bits)	15
Adaptive CB (bits)	8 + 5
Number of Pulses	3 x 2
Sign CB (bits)	1 (3 Dim.) x 2
Pulse Location (bits)	11 x 2
Gain CB (bits)	6 x 2
Total/Frame (bits)	64

sub-frame. By adopting this technique, no side information is necessary to transmit V/UV information to the decoder, because the same decision algoritnm as that in the encoder can be employed in the decoder.

4. SUBJECTIVE EVALUATION

6.4 kb/s MP-CELP speech quality is evaluated. Simulation conditions and bit allocations are summarized in Table 5.

Two kinds of experiments were designed. Experiments 1 and 2 assess performance for clear and tandem conditions and performance for background noise (car, babble and interference talker), respectively. Experiments 1 and 2 use ACR and DCR evaluation methods, respectively. Three kinds of 6.4 kb/s MP-CELP, marked as Methods 1, 2 and 3 are evaluated. Method 1 uses neither the combination search nor the adaptive pulse location restriction. Method 2 employs the combination search, but does not use the adaptive pulse location restriction. Method 3 employs both the combination search and the adaptive pulse location restriction. In Methods 2 and 3, the number of location candidates in the combination search is 16. G. 726 ADPCM at 32 kb/s and 24 kb/s, G. 729 and G.723.1 at 6.3 kb/s are

Table 6: Subjective evaluation results

	MP-CELP			G.726		G.729	G.723.1
Condition	Method 1	Method 2	Method 3	32 kb/s	24 kb/s		6.3 kb/s
Experiment 1: Clean and Tandem (ACR)							
Clean	3.01	3.40	3.48	3.32	3.07	3.72	3.50
Tandem	2.44 (2T)	2.99 (2T)	3.01 (2T)	2.95 (3T)	2.75 (2T)	-	-
Experiment 2: Background Noise (DCR)							
Car (15dB)	1.66	1.69	2.51	3.36	3.15	2.55	2.47
babb. (25dB)	1.61	1.84	2.77	3.56	3.27	3.04	2.84
Intf. (15dB)	2.22	2.72	2.86	3.13	2.76	3.28	3.01

also evaluated as reference codecs. Headphone was used as listening device. Twelve Japanese subjects took part in the experiments.

The results are shown in Table 6. From the comparison between Methods 1 and 2 in experiment 1, the combination search improves the speech quality by 0.4-0.5 in MOS values for clean and tandem conditions. In experiment 2, from the comparison between Methods 2 and 3, the adaptive pulse location restriction is effective for background noise conditions, and the speech quality is improved by about 0.9 in MOS value for car noise and babble noise conditions. The results show that the speech quality of 6.4 kb/s MP-CELP (Methods 2 and 3) speech quality is higher than that of 32 kb/s G.726 in clean speech and tandem conditions. For background noise conditions, the adaptive pulse location restriction in Method 3 significantly improves MOS value by 0.9, but its speech quality still does not reach to that of 24 kb/s G.726, except interference talker condition.

5. CONCLUSION

This paper proposes 6.4 kb/s MP-CELP speech coding with 10 ms frame. The combination search between the pulse locations and the amplitude VQ codebook improves SNR_{seg} by 3 dB, when the number of candidates is 16. The adaptive pulse location restriction improves speech quality under background noise conditions. The speech quality for 6.4 kb/s MP-CELP is higher than that of 32 kb/s G.726 and is equivalent to that for 6.3 kb/s G.723.1 with 30 ms frame in clean and tandem conditions. For background noise conditions, the adaptive pulse location restriction in Method 3 significantly improves MOS value by 0.9. The speech quality is equivalent to that for G.723.1, but does not reach to that of 24 kb/s G.726, except interference talker condition.

Acknowledgement

The authors would like to thank Mr. H. Kumagai for his help in this work.

REFERENCES

[1] ETSI/GSM, “Digital cellular telecommunications system; Enhanced full rate (EFR) speech processing transcoding (GSM 06.60),” March 1996.

[2] R. Salami, C. laflame, J-P. Adoul, A. Kataoka, S. Hayashi, C. Lamblin, D. Massaloux, S. Proust, P. Kroon and Y. Shoham, “Description of the proposed ITU-T 8 kb/s speech coding standard, ” IEEE Workshop on Speech Coding for Telecommunications, pp. 3-4, Sept. 1995.

[3] ITU-T, “Terms of reference for the variable bit rates extension to G.729,” AH12-5, Oct. 1996.

[4] K. Ozawa, S. Taumi and T. Nomura, “MP-CELP speech coding based on multi-pulse vector quantization and fast search,” IEICE Trans. Vol. J79-A, No.10, pp.1655-1663, Oct. 1996 (in Japanese).

[5] S. Taumi, K. Ozawa, M. Serizawa and T. Nomura,“Low-delay CELP with multi-pulse VQ and fast search for GSM EFR,” Proc. ICASSP, pp. 562-565, May 1996.

[6] J-H. Chen, R. V. Cox, Y-C. Lin, N. Jayant and M. J. Melchner, “A low-delay CELP coder for the CCITT 16 kb/s speech coding standard, ” IEEE J. Sel. Areas Commun. , vol. 10, no. 5 pp. 830-849, 1992.

[7] B. S. Atal and J. R. Remde,“A new model of LPC excitation for producing natural-sounding speech at low bit rates,” Proc. ICASSP, pp.614-617, 1982.

[8] K. Ozawa, S. Ono and T. Araseki, “A Study on pulse search algorithms for multi-pulse excited speech coder realization,” IEEE J. Sel. Areas Commun., SAC-4, pp.133-141, Feb. 1986.

[9] T. Nomura, K. Ozawa and M. Serizawa, “Efficient pulse search methods in CELP,” Proc. ASJ, pp. 311-312, March 1996 (in Japanese).