CELP SPEECH CODING BASED ON AN ADAPTIVE PULSE POSITION CODEBOOK

Tadashi Amada, Kimio Miseki and Masami Akamine

Kansai Research Laboratories, Toshiba Corporation,
8-6-26, Motoyama-minami-machi, Higashinada-ku, Kobe 658-0015, Japan

ABSTRACT

CELP coders using pulse codebooks for excitations such as ACELP[1] have the advantages of low complexity and high speech quality. At low bit rates, however, the decrease of pulse position candidates and the number of pulses degrades reconstructed speech quality. This paper describes a method for adaptive allocating of pulse position candidates. In the proposed method, N efficient candidates of pulse positions are selected out of all possible positions in a subframe. Amplitude envelope of an adaptive code vector is used for selecting N efficient candidates. The larger the amplitude is, the more pulse positions are assigned. Using an adaptive code vector for the adaptation, the proposed method requires no additional bits for the adaptation. Experimental results show that the proposed method increases WSNRseg by 0.3dB and MOS by 0.15.

1. INTRODUCTION

CELP coders are widely used for coding 4-8kHz bandwidth speech. Their algorithms are based on the linear predictive analysis where an LP synthesis filter and its excitation signal play an important role in producing high quality speech. In the CELP algorithm, an excitation signal consists of an adaptive code vector and a stochastic code vector. The latter has an especially strong influence on reconstructed speech quality, computational cost, and memory storage size. The stochastic code vector is the key to the performance of the CELP coding algorithm.

Pulse code vectors show good performance in terms of these factors. ACELP[1] is one of the most widely applied algorithms using pulse code vectors and is used for various codec standards [2][3][4]. However, at low bit rates of around 4kbps, a decrease in number of bits for pulse positions and signs leads to significant degradation of the speech quality.

Some approaches to reduce the degradation have been presented, such as the method in which pulse positions are restricted to even sample positions[4] and the method in which a sign is fixed for each pulse [5]. These methods keep the number of pulses the same while the number of bits for pulse positions and signs are reduced. Since pulse positions and signs are important for speech quality, they must be restricted in an efficient manner.

2. PULSE POSITION ADAPTATION

This paper describes an adaptation method for pulse position codebooks which are used in CELP coders based on pulse excitations such as ACELP. The pulse position codebooks consist of the predefined number of position data and these positions are usually fixed. In the proposed method, the positions change from subframe to subframe depending on input speech signal. The key to this adaptation is how to select the positions from the possible positions in a subframe.

The adaptation methods exploiting a pitch period have been proposed. For multi-pulse speech coding, a method in which a pitch interpolation is applied to the multi-pulse excitations [6]and a method in which multi-pulse positions are reduced based on a pitch period[7] have been proposed. For CELP speech coding, a method in which pulse position candidates are concentrated around a peak location of a pitch waveform has been proposed[8].

Using a pitch period is not the best way for the adaptation for several reasons. One reason is that a pitch period is not sufficient to represent the shape of the excitation signals. The shapes of pitch waveforms are not the same even if their pitch periods are the same. It is more effective to exploit the shape information for the adaptation. The other reason is that there is not much meaning in a pitch period during an unvoiced or a silent speech section. The adaptation method based on the pitch period cannot take voicing activity into account by itself. Another reason is that the adaptation does no work properly for a multiple pitch period or a half pitch period, which is sometimes used in the CELP coders.

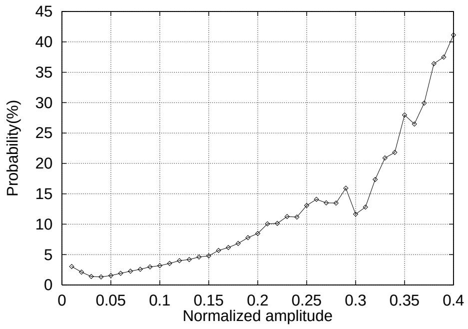

Figure 1: Probability of pulse positions

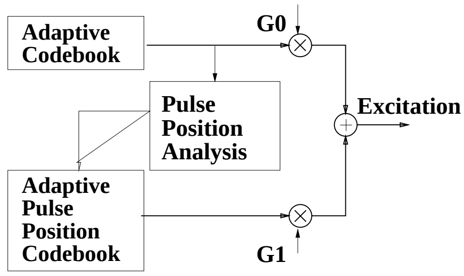

Figure 2: Block diagram of the adaptation

We propose an adaptation method using amplitude information of excitations. In order to investigate the relationship between pulse positions and amplitude of the excitations, an experiment was performed. The codec used in this experiment has a 3-pulse ACELP structure for producing a stochastic excitation. Position codebooks for each pulse include all samples in a subframe. This excitation could be considered to be an optimal excitation for 3-pulse ACELP. Comparing the pulse positions of the excitation with the optimal stochastic excitation, which is produced by inverse filtering of a target speech vector, a relationship between pulse positions and amplitude of optimal stochastic excitations was found.

Figure 1 shows the relationship. The vertical axis shows the probability that an optimal pulse is settled. The horizontal axis shows the smoothed amplitude of optimal stochastic excitations normalized by their energy. From this experiment, it can be seen that pulses are to be placed for the sample where the smoothed amplitude of the optimal stochastic excitation is large. We exploit this relationship in the adaptation of pulse position candidates. In the proposed method, N efficient candidates of pulse positions are selected out of all possible positions in a subframe. The larger the amplitude of an optimal stochastic excitation is, the more positions are assigned. Although the number of pulse position candidates is constant in a subframe, the efficient allocation of the pulse position candidates improves the codec performance.

In this paper, in order to obtain the smoothed amplitude, an adaptive code vector is used instead of the optimal stochastic excitation. One reason is that envelope of the adaptive code vector is similar to that of the optimal stochastic excitation, especially in voiced speech sections. Another reason is that it can be used in both an encoder and a decoder without additional information.

Figure 2 shows the block diagram of the proposed method. The flow of the adaptation is as follows: (1) an adaptive code vector is selected from the adaptive codebook, (2) smoothed absolute amplitude of the adaptive code vector is obtained in the pulse position analysis, (3) depending on the smoothed absolute amplitude, predefined number of pulse positions is allocated so that the larger the amplitude envelope is, the more pulse positions are assigned.

The advantage of using amplitude information is explained in Figure 3. In this figure, amplitude envelopes of the adaptive code vectors are shown. Small arrows indicate pulse position candidates obtained by the proposed method. Large arrows indicate a peak location of a pitch waveform.

In Figure 3(a), there are 2 pitch periods in a subframe and each pitch waveform has two envelope peaks which are shown as “A” and “B” in the figure. Since the proposed adaptation uses amplitude information, pulse position candidates are concentrated not only in section B, but also in section A. In the conventional method using a pitch period, pulse position candidates are concentrated only in section B where the peak location is placed.

Figure 3(b) is an example of a subframe which has weak periodicity, such as in an unvoiced subframe. In this case, the strength of the adaptation is weak and a fairly uniform allocation is obtained. The proposed method inherently has the ability to control the strength of adaptation, depending on the strength of the periodicity of an adaptive code vector.

Figure 3(c) is an example of a subframe where pitch doubling occurs. The proposed method is not influenced by the incorrect pitch period because it exploits the amplitude information. The conventional method, however, cannot allocate pulse position candidates based on a true pitch period.

Figure 3(d) is another example which shows the benefit of using amplitude information. In these sub-

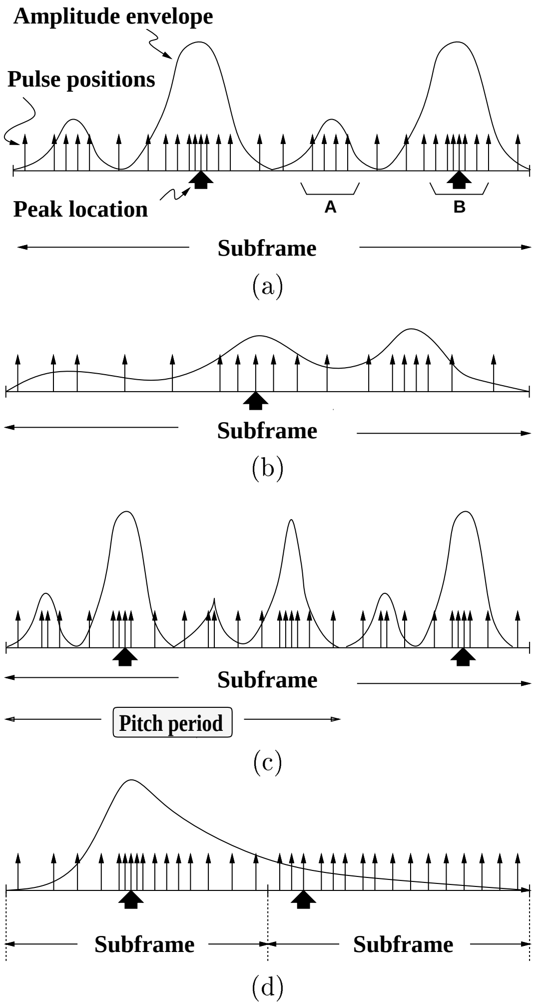

Figure 3: Adaptation of pulse position candidates: (a)strong periodicity;(b)weak periodicity;(c)multiple pitch period; (d)pitch period longer than a subframe

frames, the pitch period is longer than the subframe length. In the first subframe, pulse positions are concentrated around the peak. In the second subframe there is no peak of the pitch waveform so that pulse positions are allocated fairly uniformly. The conventional method cannot make use of the difference between the first subframe and the second subframe.

From these examples, it can be seen that using amplitude information is effective for adaptive allocating of pulse positions.

3. SIMULATION RESULTS

A CELP codec used in this simulation has 3-pulse ACELP excitations. Pulse position candidates of each channel were adaptively changed using the proposed method. Table 1 shows how N pulse position candidates $(S_1, S_2, \cdots, S_N)$ were distributed to 3 channels.

The bit rate of this codec was 3.7kbps and 12 bits/subframe were assigned for ACELP excitations. The adaptive codebook had 3 modes: absolute lag mode, differential lag mode and fixed-codebook mode.

Around 4kbps, the number of pulses in a subframe tends to be low so that reconstructed speech has pulse-like noise. To reduce this noise, a pulse shaping filter is used. It is a 20th order FIR filter whose coefficients are changed depending on the mode for the adaptive codebook. A closed learning method was applied to obtain the coefficients and they are shown in Figure 4.

In Table 2, average WSNRseg (perceptually weighted segmental SNR) for 12 male sentences and 12 female sentences are shown. WSNRseg means segmental WSNR

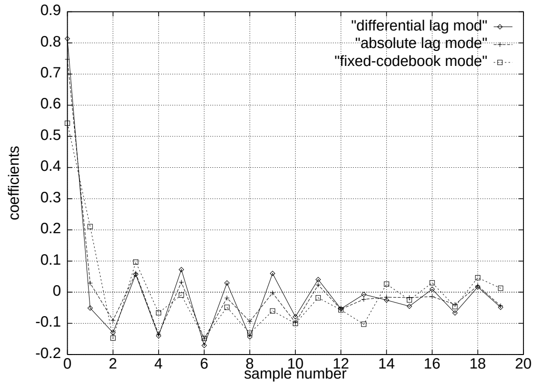

Figure 4: Pulse Shaping Filter

Table 1: Pulse position assignment

Channel No.	positions			
1	S_1	S_4	$\cdots$	S_{N-2}
2	S_2	S_5	$\cdots$	S_{N-1}
3	S_3	S_6	$\cdots$	S_N

Table 2: Simulation result in WSNRseg

	without adaptation (dB)	with adaptation (dB)
Female	5.17	5.36
Male	4.23	4.60
Total	4.70	4.98

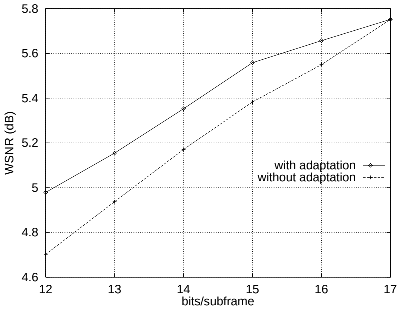

Figure 5: Bit rates versus WSNRseg

which is defined as

$$WSNR = 10\log_{10}\frac{\sum_{i=1}^{M} s(i)^2}{\sum_{i=1}^{M}(s(i)-\hat{s}(i))^2} \quad (1)$$

where $s(i)$ is a perceptually weighted input speech signal, $\hat{s(i)}$ is a perceptually weighted reconstructed speech signal and M is the number of samples in the signal. Since the codebook search is performed under the perceptually weighted speech level, WSNRseg is better than SNRseg for evaluating the excitation performance. "Without adaptation" means that 24 pulse position candidates (8 candidates × 3 channels) were uniformly distributed in a subframe (80 sample). This method is similar to the conventional one which restricts pulse positions to the even sample positions[4]. "With adaptation" means 24 candidates were adaptively distributed by using the proposed method. Both methods used the same pulse shaping filter. Using the proposed method, WSNRseg increases by 0.3dB.

Figure 5 shows the relationship between WSNRseg and bits for pulse excitation under the conditions of "with adaptation" and "without adaptation". The number of bits for pulse excitation varied from 12 to 17 and the number of bits for other parameters was the same. From Figure 5, it can be seen that the lower the number of bits for excitation is, the more effective the proposed method is. The proposed method shows better performance at low bit rates.

4. LISTENING TEST RESULTS

A subjective test has been performed. 8 sentences were rated by 8 listeners. Absolute category rating (ACR) was used and each sentence was scored from 1(bad) to 5(excellent) by listeners. The codec used in this experiment was a 3.7kbps ACELP codec which is described in Section 3. The result (MOS) is shown in Table 3. The codec using proposed adaptation shows higher MOS than the codec without adaptation in all conditions. The proposed adaptation increases average MOS by 0.15.

Table 3: Subjective test result in MOS

	without adaptation	with adaptation
Female	2.94	3.06
Male	2.72	2.91
Total	2.83	2.98

5. SUMMARY

A new adaptation method for the pulse position codebook was proposed. The method changes positions in the pulse position codebooks depending on the amplitude envelope of an adaptive code vector. Computer simulations showed that the proposed method increases WSNRseg by 0.3dB. Subjective tests showed that the proposed method increases MOS by 0.15 point. The proposed method makes it possible for CELP coders using pulse excitations such as ACELP to achieve high quality speech at low bit rates of around 4kbps. Improving an error resilience is a future study.

6. REFERENCES

[1] J.P. Adoul, P. Mabilleau, M. Delprat and S. Morissette, "Fast CELP Coding Based on Algebraic Codes," IEEE Proc. ICASSP'87, pp.937-940, 1987.

[2] TIA IS-127, "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems," Sept. 1996.

[3] ITU-T Recommendation G.729, "Coding of Speech at 8kbit/s using Conjugate-Structure Algebraic-Codec-Excited Linear-Prediction (CS-ACELP)," 1995.

[4] ITU-T Recommendation G.723, "Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 & 6.3 kbit/s," 1995.

[5] A. Kataoka, S. Kurihara, and S. Hayashi, "A 6.4-kbit/s Extension to the G.729 Speech Coder," IEEE Speech Coding Workshop, pp.69-70, 1997.

[6] K. Ozawa and T. Araseki, "Low Bit Rate Multi-pulse Speech Coder with Natural Speech Quality," IEEE Proc. ICASSP'86, pp.457-460, 1986.

[7] S. Hyashi, M. Suguimoto and G. Kishi, "Adaptively controlled 4.8 kbit/s speech coders with low distortion," Electron. Commun. Jpn. 1, Commun., vol.76, no.10, pp.26-33, 1993.

[8] H. Ehara, K. Yoshida and T.Yagi, "A Study on Phase Adaptive Pulse-Search in CELP Coding," Proc. Fall Meeting of ASJ, pp.273-274, Sept. 1996. (in Japanese)