A 4 KBIT/S RENEWAL CODE EXCITED LINEAR PREDICTION SPEECH CODER

Hong Kook Kim[1] *Yong Duk Cho*[2] *Moo Young Kim*[3] *Sang Ryong Kim*[4]

Human & Computer Interaction Lab.
Information Processing Sector
Samsung Advanced Institute of Technology
San 14, Nongseo-Ri, Kiheung-Eup, Yongin-City, Kyungki-Do 449-712, Korea
{kimhk[1] ydcho[2] moo[3] srkim[4] }@{saitgw[1,2,4] blue[3]}.sait.samsung.co.kr

ABSTRACT

This paper proposes a new 4 kbit/s speech coder based on CELP structure with 45 ms total codec delay. The coder is mainly featured by the renewal codebook of the excitation signal and the linked split-vector quantizer of LSPs which enable the coder to get high quality speech at low bit rate. In addition, techniques of the formant enhancement in spectral envelop and the harmonic recovery in transient region are also introduced to reduce buzzy and hoarse sounds, respectively. From the intensive listening test with intermediated response system (IRS) speech, we obtained the comparable subjective quality to 32 kbit/s ADPCM (ITU Recommendation G.726) under nominal speech input level of -26 dB overload.

1. INTRODUCTION

During the past two decades, ITU-T has standardized several speech coders such as A/μ-law PCM at 64 kbit/s, ADPCM at 16, 24, 32 and 40 kbit/s, LD-CELP at 16 kbit/s, and CS-ACELP at 8 kbit/s. Now the next speech coding standardization at 4 kbit/s is under processing [1]. Around 8 or 16 kbit/s, CELP algorithm reproduces high quality speech comparable to 32 kbit/s ADPCM. But at low bit rate, 4 kbit/s, the quality of the reconstructed speech of the CELP may deteriorate rapidly for the lack of bits representing excitation sequence faithfully.

To overcome the problem in the conventional CELP, we propose a new CELP coder having a novel excitation scheme. In the proposed renewal CELP (RCELP), the excitation signal is generated from adaptive codebook, so high quality speech can be obtained with lower bits. Also, to improve the performance of the RCELP coder, it includes additional features. They are composed of the formant enhancement to diminish buzzy sound, the efficient LPC quantization to preserve spectral transparency, and the harmonic structure reproduction at transient region to reduce hoarse sound.

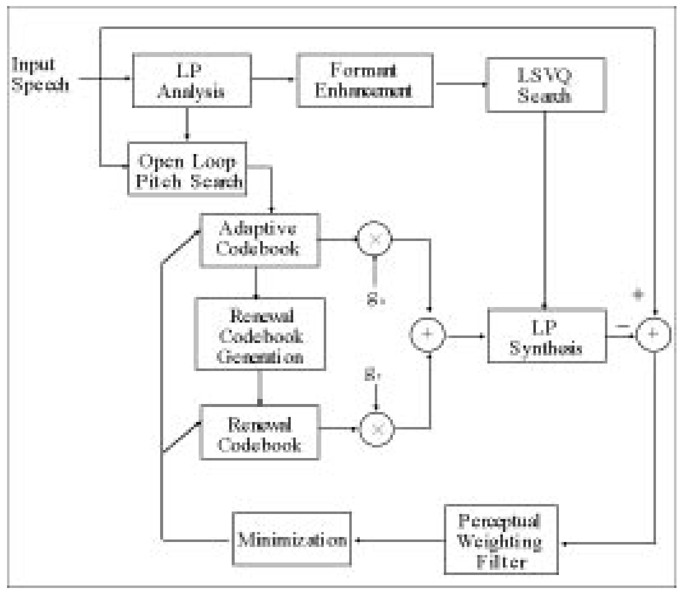

Figure 1. RCELP encoder structure.

2. SPEECH CODER STRUCTURE

The encoder of RCELP mainly consists of LPC analysis, formant enhancement, linked split-vector quantizer (LSVQ), open-loop pitch search, adaptive codebook search, renewal codebook generation and search, and perceptual weighting filter as shown in Figure 1. The decoder of RCELP can be realized as a part of the encoder.

2.1. LP Analysis and Formant Enhancement

The short-term analysis is done by the 10-th order linear prediction (LP) filter. A frame is split into three subframes and each subframe is further divided into two parts (subframe-parts). The LP analysis is carried out for each part of subframes using the autocorrelation method with an asymmetric window. The maximal weighting point of the window with a length of 240 is positioned at the center

Table 1. The performance comparison between SVQ and LSVQ

Quantizer	bits/frame	SD (dB)	2 - 4 dB (%)	> 4 dB (%)
SVQ	24	1.03	1.60	0.12
LSVQ	23	1.01	1.53	0.00

of the corresponding subframe-part. A spectral smoothing technique is applied to widen the spectral peak with binomial window of 90Hz. It is well known that spectral smoothing technique is worth extracting the LPC coefficients well in high pitch voice of female. For the lower pitched speech, we can sometimes feel the buzzy sound due to applying this technique. In order to prevent this phenomenon selectively, we perform the formant enhancement technique to LPC coefficients. The formants are enhanced by radially shifting the poles of the LPC synthesis filter into the unit circle, and thus, we can improve the quality of RCELP by eliminating this buzzy sound.

2.2. Spectral Quantization

LP coefficients derived from the last subframe of the current frame are converted into line spectrum pair parameters (LSPs). These LSPs are quantized by using the LSVQ with 23 bits/frame [2]. Table 1 shows that LSVQ at 23 bits/frame outperforms split VQ (SVQ) [3] at 24 bits/frame in views of spectral distortion (SD) and the percentage of outliers. So we can get 1 bit reduction over SVQ while preserving spectral transparency. The quantized LSPs are linearly interpolated with those of the previous frame to construct the LPC synthesis filter of each subframe-part for both the encoding and decoding process.

2.3. Open-loop Pitch and Adaptive Codebook Search

Three tap adaptive codebook approach is implemented as long-term predictive analysis. As a starting point to reduce the search space of the codebook lag in adaptive codebook search, we use the open-loop pitch of current frame. Searching an open-loop pitch is based on the modified LPC residual excitation method. First, we modify the LPC residual by applying a window function to emphasize the second subframe. We extract the open-loop pitch of the second subframe, and then choose each of the best adaptive codebook lag for the second subframe by searching several lags around the extracted pitch. In the same manner, the best candidate lag of the third subframe can be chosen. On the other hand, the adaptive codebook lag of the first subframe is decided from the open-loop pitch of previous frame. Performance of the proposed procedure is better in recovering the harmonic structure of transient region, while being same to that of the conventional adaptive codebook search in steady state region.

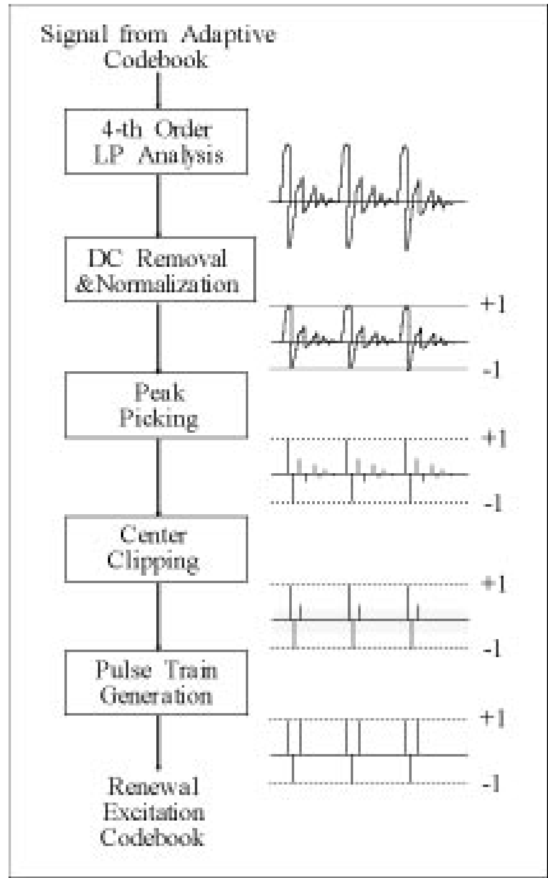

Figure 2. Renewal codebook generation procedure.

The pitch lags are encoded with 7 bits for the second subframe and 5 bits for the first and third subframes. The gains of adaptive codebook for each subframe are coded by the LBG vector quantizer [4] whose size is 64.

2.4. Renewal Excitation Codebook Generation and Search

To efficiently represent the excitation signal in CELP, lots of schemes have been proposed such as a random codebook, an algebraic structured codebook, and so on [5]. Generally, lowering the bits to represent the excitation signal causes the performance to degrade rapidly. So we propose a novel excitation codebook to represent the excitation signal efficiently, named renewal excitation codebook.

Table 2. **Bit allocation of 4 kbit/s RCELP.**

Parameter		Assigned Bits			
		SF 1	SF 2	SF 3	Frame
LSP		23			23
Adaptive Codebook	Pitch	2 (or 3)	7	3 (or 2)	12
	Gain	6	6	6	18
Renewal Codebook	Index	5	5	5	15
	Gain	4	4	4	12
Total		80			

The proposed renewal excitation codebook is not fixed but varied according to the adaptive codebook content. To generate a renewal excitation sequence from the adaptive codebook signal, the 4-th order LP analysis followed by D-C removal and normalization of the LPC residual, peak picking, center clipping, and pulse train generation are applied to the adaptive codebook signal as shown in Figure 2. The LP analysis eliminates the short-term correlation of the adaptive codebook sequence. Subsequently, by the non-linear processings such as the peak picking and the center clipping, we can force the resultant renewal code sequence to contain high frequency information. By overlapping and shifting the renewal excitation sequence, 32 entries of renewal excitation codebook are obtained. The gain of the excitation sequence for each subframe is quantized with 4-bit non-uniform scalar quantizer.

2.5. Bit Allocation of 4 kbit/s RCELP

The speech signal is stored and processed frame by frame with a frame length of 20ms. The lookahead of 5ms is needed for the short-term predictive analysis in the encoder. Assuming the processing delay is equal to the frame length, the total codec delay becomes 45ms. Each frame is divided into three subframes with a length of 53, 53, and 54 samples, respectively. Table 2 shows the bit allocation of the 4 kbit/s RCELP codec.

3. PERFORMANCE OF RENEWAL EXCITATION CODEBOOK

For the efficient representation of excitation sequence we have investigated lots of experiments. From the experiments we found that a speech coder using random excitation codebook with 32 code entries results in poorly reconstructed speech. It is due to the lack of excitation code sequences. In the case of algebraic excitation [6], we could obtain high quality speech at the expense of 4 pulses/5 ms. It is, however, expected that the number of pulses should be reduced into 2 pulses/6.7 ms to obtain a speech coder operating at 4 kbit/s. But this also resulted in seriously degraded speech quality. Figure 3(a)-(c) show original input speech, desired excitation signal to be modeled, renewal excitation, and synthesized speech by renewal excitation, respectively. This figure shows that the renewal excitation codebook can generate a reasonable excitation sequence and thus we can reconstruct speech of high quality.

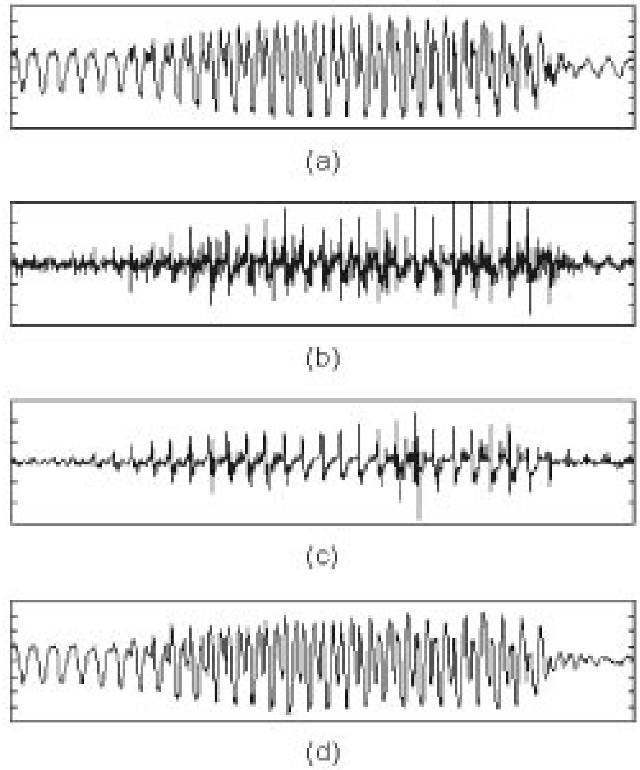

Figure 3. The comparison of waveforms between (a) original input speech, (b) desired excitation sequence, (c) renewal excitation sequence, and (d) synthesized speech from renewal excitation. Waveforms of (b) and (c) are amplified so that their magnitudes are adjusted to those of waveforms of (a) and (d).

4. RCELP EVALUATION

4.1. Test Environment

Overall test procedure was performed by following the ITU-T temporary document 83 of Study Group 15 [7]. The data for the qualification test were prepared from pre-recorded CD-ROM material of NTT Advanced Technology Co. (NATC) which follows the recording guideline of

Table 3. The result of subjective speech quality test between 32 kbit/s G.726 and 4 kbit/s RCELP. $\overline{MOS_c}$ and S_c mean average and standard deviation, respectively, over four speakers, M1, M2, F1, and F2.

Codec	MOS_c				$\overline{MOS_c}$	S_c
	Male Speaker		Female Speaker			
	M1	M2	F1	F2	All	All
G.726	3.75	3.69	3.69	3.85	3.75	0.46
RCELP	3.77	3.67	3.54	3.67	3.66	0.57

ITU-T Recommendation P.80 [8]. The test was done with eight Korean sentences pronounced by four Korean speakers. Each sentence was processed by both RCELP and G.726 coders, and listened by twenty-four Koreans for MOS test.

4.2. Test Results

From the subjective speech quality test, 4 kbit/s RCELP and 32 kbit/s G.726 scored 3.67 and 3.75 in MOS, respectively, as shown in Table 3. This result indicates that the subjective quality of RCELP is not worse than that of 32 kbit/s G.726 under the error free channel and quiet environment [9]. However, other listening tests under channel errors or noise environments show seriously degraded quality of RCELP in comparison with that of 32 kbit/s G.726.

5. CONCLUSION

We developed a high quality speech coding method operating at 4 kbit/s called RCELP. The coder is mainly characterized by the renewal excitation codebook and the linked split-vector quantizer of LSPs. The former generates diverse excitation sequences with small number of bits and the latter transparently quantizes spectral envelop at 23 bits/frame. In addition, the decoded speech is spectrally enhanced by emphasizing the formants of LPC spectral envelop and recovering harmonic structure efficiently at transient region. From intensive listening test, it is shown that this coder reproduces high quality speech at 4 kbit/s by virtue of the above characteristics and enhancement techniques.

REFERENCES

[1] R. V. Cox, "Speech Coding Standards," in *Speech Coding and Synthesis,* W. B. Kleijn and K. K. Paliwal, Eds., Amsterdam, Elsevier Science, pp. 49-78, 1995.

[2] M. Y. Kim, N. K. Ha, and S. R. Kim, "Linked Split Vector Quantization of LPC Parameters," in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing,* Atlanta, GA, Vol. II, pp. 741-744, Apr. 1996.

[3] K. K. Paliwal and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *IEEE Trans. Speech and Audio Process.,* Vol. 1, No. 1, pp. 3-14, Jan. 1993.

[4] Y. Linde, A. Buzo, and M. R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Commun.,* Vol. 28, No. 1, pp. 84-95, Jan. 1980.

[5] A. Gersho, "Advances in Speech and Audio Compression," *Proc. of the IEEE,* Vol. 82, No. 6, pp. 900-918, June 1994.

[6] R. Salami, C. Laflamme, J-P. Adoul, and D. Massaloux, "Toll Quality 8 Kb/s Speech Codec for the Personal Communications System (PCS)," *IEEE Trans. Veh. Technol.,* Vol. 43, No. 3, pp. 808-816, Aug. 1994.

[7] ITU-T Temporary Document TD.83 (WP 2/15), "Subjective Qualification Test Plan for the ITU-T 4 kbit/s Speech Coding Algorithm," ITU-T SG15, Nov. 1995.

[8] ITU-T Recommendation P.80, "Methods for Subjective Determination of Transmission Quality," ITU-T SG12, Mar. 1993.

[9] ITU-T Delayed Contribution D.827 (WP 2/15), "Qualification Test of Samsung's Renewal CELP (RCELP) for the ITU-T 4 kbit/s Speech Coding Standard," ITU-T SG15, May/June 1996.