On a Reduction of Pitch Searching Time by Preliminary Pitch in the CELP Vocoder

Seonggyun BAE, Hyungrae KIM

Department of Electronic Engineering
Kon-kuk University
93-1, Mojin-dong seongdong-ku
Seoul 133-701
Republic of Korea

Daesik KIM, Myungjin BAE

Department of Telecommunication Engineering
SoongSil University
1-1, Sangdo-Dong, Dongjak-Ku
Seoul 156-743
Republic of Korea

ABSTRACT Code Excited Linear Prediction(CELP) as a speech coder exhibits good performance at data rates below 4.8 kbps. The major drawback to CELP type coders is their large amount of computation. In this paper, we propose a new pitch search method that preserves the quality of the CELP vocoder with reduced complexity. The basic idea is to restrict the pitch searching range by estimating the preliminary pitches. Applying the proposed method to the CELP vocoder, we can get approximately 87 % complexity reduction in the pitch search.

1. INTRODUCTION

The main methods of coding for transmission or storage of speech signals to the memory can be generally classified into following three types; waveform coding, source coding, and hybrid coding method.

In the waveform coding methods, the repetitive unnecessary redundancies in speech waveforms are removed before it is transmitted through the transmission channel or stored in some storage medium. These types are PCM, ADM, ADPCM, etc. In recent, because of the improvement of the manufacturing techniques and the development of the DSP(Digital Signal Processor) algorithms, the standard ADPCM chip has realized with a bit rate of 32 kbps. Also, the waveform coding methods can maintain the high quality and personality, because, in the processing procedure, both the vocal tract filter informations which represent the meaning of message and the excitation informations which reflect the personality and the feeling of a person are transmitted without being seperated into two parts.

The methods of source coding are based on the speech production model. In speech signals these methods separate the excitation information and the filter information, and then each is coded. The methods that belong to this category are LPC, PARCOR, LSP, MBE, formant coding, etc. These algorithms are very efficient in memory capacity because these have a low transmission rate of 1 kbps.

The hybrid coding methods have the memory efficiency of source coding and

the naturalness and intelligibility of waveform coding. In these methods, the formant information is encoded generally by Linear Predictive Coding method(LPC), and according to how to encode the residual signal, these methods are classified into RELP, VSELP, MPLP, and CELP. Among these methods, recently Code Excited Linear Prediction(CELP) is adopted for the mobile communication.

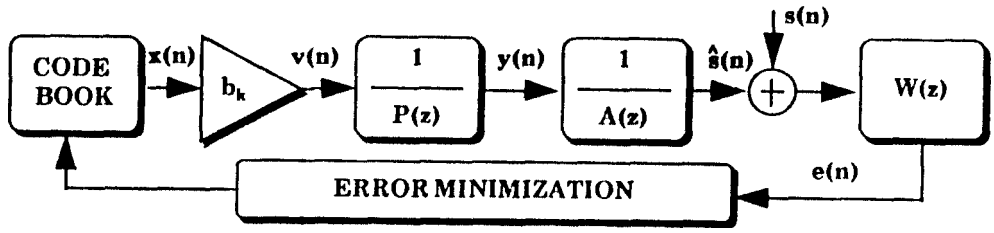
The CELP coding encodes the pitch period of speech signal by applying the pitch filter. The pitch searching method applied primarily to this pitch filter is the correlation method using the pitch lag. The pitch lag and the gain of the pitch filter in pitch searching method by correlation is decided optimal correlation value by searching the correlation of all pitch lags being pitches with two signals. But this pitch searching procedure must examine about all pitchs within the intervals, therefore it is difficult to implementation with DSP chip and has many handling time.

For that reason, in this paper, we propose a new method to reduce the pitch searching time by preliminary pitch in the CELP vocoder.

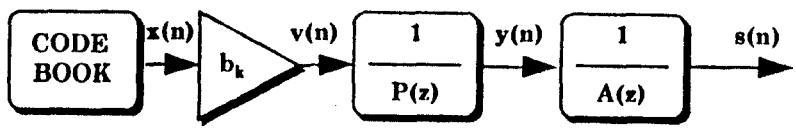
2. THE PRINCIPLE OF CELP VOCODER

The block diagram of CELP vocoder is shown as Fig. 2-1. The formant synthesis filter usually be applied to the 10th order LPC coefficients of all pole structure. When being encoded, the LPC coefficients are converted to the LSP coefficients to reduce the distortions from quantizing errors. And then they are recovered to the LPC coefficients when they are decoded in the decoder. The LPC coefficients are encoded every a frame of 20 ms, and provided differently each subframe of 5 ms after interpolating. Also the excited source parameters are converted newly every subframe of 5 ms.

The encoder and decoder make use of each two excited source in the CELP. First, the excited source is long-term(pitch) predictive state or adaptive codebook. Second, the excited source is taken in excited codebook.



A) A CELP speech encoder



B) A CELP speech decoder

Fig. 2-1. A CELP speech coder

The length of codeword is 128 samples in the low transmission rate. These two excited sources are multiplied by conformational gain term to those, and then the results are summed. These are a combined excitation sequence. The excited output of each subframe is applied to change the long-term filter state of the adaptive codebook to be made use of in the next subframe.

In CELP vocoder, because of applying vector quantization to the residual signal which the formant information is strained, the data of the residual signal component to be transmitted is the index of codebook. Consequently, the transmission rate can be lowered below 4.8 kbps and if these parameters are transmitted together with additional error correcting code, it is robust coding method under transmission noise. For it is analyzed repeatedly to maintain optimal talk quality with applying analysis by synthesis method, the quality is excellent in the low transmission rate.

The CELP vocoder has complex structure as it must always compare the synthesized signal with original signal. Specially, it is required a lot of computation time when coding, and the most of computation time is caused by finding the input excitation signal and the coefficient of pitch filter p(z).

The pitch searching is to obtain pitch period information corresponding to long term correlation of speech signal. The pitch analysis of speech signal sampled in 8 kHz is performed every 5 ms. The spectrum analysis is fulfilled with open circuit structure, but the pitch analysis must be done with closed one. That is, the value satisfying optimally pitch delay condition is determined through repetitive comparison. The pitch gain is quantized relative gain by using resulted pitch delay. It is important that this pitch searching procedure has a large effect on computation time of CELP vocoder together with codebook search.

3. PITCH SEARCHING METHOD

The pitch searching procedure is to decide the value satisfying optimally pitch delay condition with a repetitive comparison between original and synthesized speech signal. That is, this procedure is that the correlation values achieve with altering gradually time delay about original speech signal and that the time delay which has the maximum value among these values is detected by pitch period.

So far the proposed methods to improve pitch search are self-excited structure[7], expanded adaptive codebook structure[5], delta pitch search structure[6], etc. These are methods reducing pitch search time by using correlation between adjacent pitch periods when usually searching pitch period.

In searching pitch, according to time delay, the correlation value E(.) of the residual signal s(n) is computed as follow:

$$E(L)=\frac{\sum_{n=0}^{M-1}(s(n)s(n-L))}{\sum_{n=0}^{M-1}(s(n-L)s(n-L))} \qquad (3\text{-}1)$$

where M is subframe length and L is time delay. Therefore, the calculated correlation is obtained the value near 100 % during each pitch period, and the similarity differs according to amplitude variation and periodicity of waveform. And when the time delay conforms to the constant times of periodicity in speech waveform, the correlation has maximum value.

To obtain most desirable time delay in pitch search, the correlation equation as Eq.(3-1) must be repeatedly performed about all pitch delays as much as possible. This is required many computation time owing to perform the multiplication and addition of each M times in every time delay L(from 20 to 147). Therefore, the pitch searching time of CELP vocoder needs over 5 MIPS, when implementing with the latest DSP chip, and this computation time is occupied the half of the implementation time in CELP vocoder. For that reason, we need the new algorithm to reduce only the pitch search computation time without pitch search errors.

4. REDUCTION OF PITCH SEARCHING TIME

The pitch searching is to obtain the pitch gain and the pitch lag when the synthesized speech signal composed of residual signal is almost similar to original speech[1][10][11]. And the correlation according to time delay is searched for maximum correlation. To obtain the time lag which has maximum correlation, it needs to search all durations being pitches sequentially. For this reason, the sequential pitch searching method retains too much of the time in processing. To reduce computation time in processing, we must know the duration of high correlation in pre-processing. In this way, by searching pitch only during durations, the computation time can be reduced.

The pitch in speech signals is defined the duration between peaks or valleys. In the case of pitch detection using the peaks, the correlation is high at the time lags being prominent peaks. The other way, using the valleys, the correlation is high at the time lags being prominent valleys.

Also, the period of pitch is not existed in 2.5ms, the preliminary pitch which can be applied to pitch search procedure is obtained by decimating waveform as follows:

First, a frame of 19 samples is invested with duration number i. At this time, with computing the maximum peak of ith composed 19 samples, the magnitude and the position of it are stored in peak buffer p(i, 1) and p(i, 0) respectively. Likewise, with measuring the minimum valley, the magnitude and the position of it are stored in valley buffer v(i, 1) and v(i, 0) separately.

In this way, the peak and the valley are found, the preliminary pitch may have error of a few sample because of the effect on the phase variation of the third formant in speech signal, therefore the errors of the higher formant can be removed by performing above decimation procedure after speech signal is filtered by Hanning;

$$s'(n-2) = \frac{s(n)+2s(n-1)+3s(n-2)+2s(n-3)+s(n-4)}{9} \qquad (4-1)$$

where, the cutoff frequency of this filter is 2.67 kHz. To use the detected peak and valley as preliminary pitch, if the difference between the first founded prominent peak(valley) and the next peak(valley) exist only in interval as following, the autocorrelation as Eq. (3-1) must be performed ;

$$T_p(2i) = p(i,0) - T_{hp} \quad \text{and}$$

$$T_v(2i+1) = v(i,0) - T_{hv}, \quad i = 1, 2, \ldots, 12 \qquad (4-2)$$

where T_{hp} is the position of the first prominent peak and T_{hv} is the position of the first valley.

The collection of detected preliminary pitch is applied to Eq.(3-1), and then the value of pitch filter L is detected by the $T_p(i)$ composed maximum $E(T_p(i))$, and the coefficient of pitch filter is

$$b_i = E_{xy}/E_{yy}$$

$$= \frac{\sum_{n=0}^{M-1}(s(n)s(n-L))}{\sum_{n=0}^{M-1}(s(n-L)s(n-L))}. \qquad (4-3)$$

The peak and valley are searched one per 19 samples by considering separately the interval of peaks and valleys. And if the interval of preliminary pitch is found each, the pitch searching time is reduced much more than that of the full pitch search method as following;

$$T_R = \frac{2}{19} \times 105 = 11\ \%. \qquad (4-4)$$

where T_R is a required time rate in computation and the adding 5% in computation time is considered the time that performs decimation to find the preliminary pitch.

5. EXPERIMENT & RESULTS

For the simulation, we used the IBM-PC/486DX II(50) interfaced with A/D converter for input and output of speech signals. The sampling frequency is 8 KHz and the quantization level is 12 bit/samples. And, the speech data composed of 3 Korean speaker's utterances(a female 20 years old, a male 22 years old, and a male 28 years old) and the following sentences were spoken each 5 times.

Utterance 1) / IN SOO NE KO MA GA CHUN JAE SO NYUN WL JO A HAN DA /

Utterance 2) / JE SU NIM KE SEO CHUN JI CHANG JO WI KIO HUN WL MAL SUM HA SEOSS DA /
Utterance 3) / SOONG SIL DAE JUNG BO TONG SIN GONG HAK KWA UM SEONG SIN HO CHU RI YUN GU SIL /
Utterance 4) / GONG IL I SAM SA O RUK CHIL PAL GU /

Where the meaning of utterance 1 is "Insoo's young boy likes a genius kid", utterance 2 is "Jesus spoke of the lessons that the creation of the heavens and the earth", utterance 3 is "Speech signal processing team at the department of information and telecommunication, Soongsil University", and utterance 4 is "one two three four five six seven eight nine", spoken in Korea.

The proposed pitch searching in CELP vocoder is performed with the C-language. In the block diagram of Fig. 5-1, the block of proposed preliminary pitch detection procedure is shown as a dotted line in pre-processing. Where the 1/A(z) is the transfer function of formant filter, the A(z)/A(z/a) is the response of perceptual weighting filter, and the ZIR is zero input filter response in previous state. The $y_L(n)$ is the synthesized speech with the pitch lag L, the E_{xy} is the cross-correlation between the input speech and the synthesized speech, and the E_{yy} is the autocorrelation of synthesized speech signal.

For performance test of the pitch search method, the procedure of simulation is divided into two part. First, the sequential pitch search method is executed to pitch searching by increment the pitch lag L during period of pitch searching(from 20 to 147). The result shows in Fig.5-2(e). In this case the distinct correlation is obtained each pitch period.

The second part of processing is implemented by proposed method. After the speech signal passes through Hanning filter that the cut-off frequency is 2.67 kHz, we detect a peak and valley per 19 samples in a frame. And, we obtained the time difference between the first prominent peak and the next peak, if the obtained time interval is within from 20 to 147 samples, it is considered as the preliminary pitch. Also, similarly the preliminary pitch is extracted at the detected valleys. The results show in Fig. 5-2(b) and 5-2(c). Simultaneously the pitch search is not performed in period which has not preliminary pitch. The pitch is determined by interval that has maximum prediction gain among the preliminary pitches. The correlation value sets zero in these skip duration. This result shows Fig. 5-2(d). In this result, the position of the maximum correlation peak is correctly the same that pitch.

To obtain the difference time of the pitch search in two procedure, the average searching time of 1 sec unit is obtained for above utterances. The sequential pitch search method needs the average 7.52 sec, but the proposed method needs the average 1.02 sec. The pitch search time is reduced about 87 % as relative computation time. But, according to computer types, the estimated time is different, we considered only relative reduction rate in time evaluation. And, the result of performing pitch search was degenerated average 0.75 dB(from average 10.89 dB to 11.64 dB) than that of the the sequential pitch search, but the pitch searching time was reduced about 87 % as relative computation time.

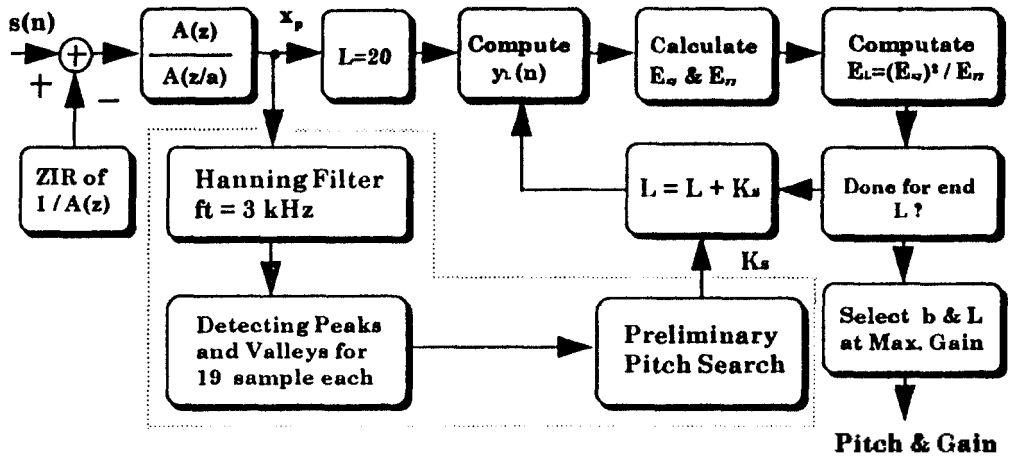


Fig. 5-1 The pitch search algorithm proposed in this paper.

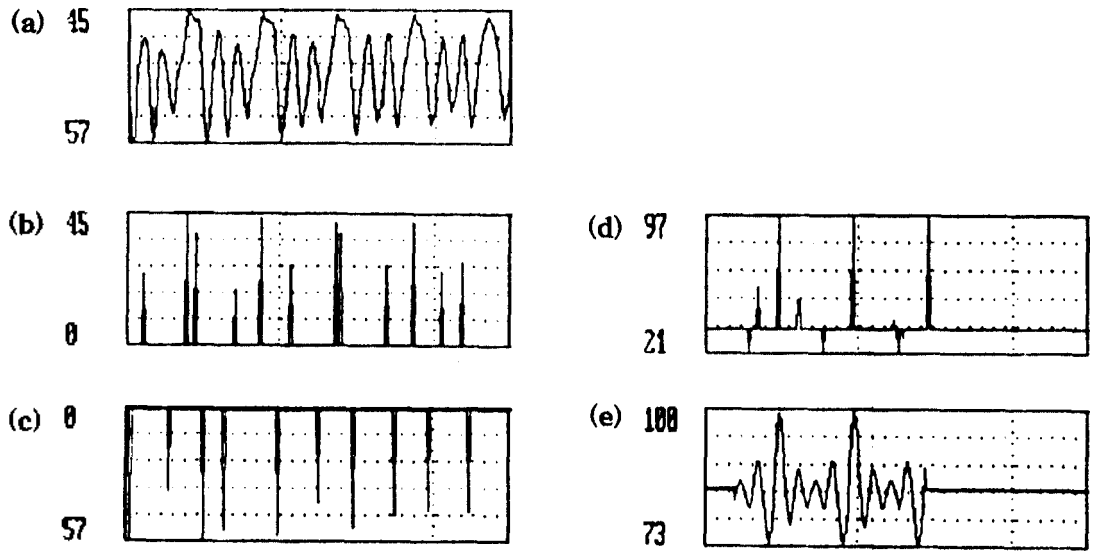


Fig. 5-2 A result for utterance 1.
(a) Speech Signal (b) Waveform Decimated for Peaks
(c) Waveform Decimated for Valleys
(d) Pitch Filter Coefficients at Preliminary Pitches
(e) Pitch Filter Coefficients by Full Search

6. CONCLUSION

The CELP vocoder obtains high talk quality by using analysis by synthesis that compares input speech signal with synthesized. But it is difficult to implement in real time with the existing DSP chip because the computation time is very large. In CELP vocoder, the pitch searching time hold the half of coding time. Accordingly, we proposed a new algorithm to reduce the pitch searching time by using preliminary pitch in the CELP vocoder.

The pitch period is the interval between peaks or valleys in speech waveform.

Additionally, the pitch in speech processing is detected generally above 2.5 ms. Thus, by using these properties, we detected the prominent peaks(valleys) in 2.375 ms before detecting pitch, and then used these interval as preliminary pitch. As results, the pitch searching time is improved by detecting the coefficient of pitch filter about preliminary pitches. With this proposed algorithm, the result of performing pitch search was degenerated average 0.75 dB than that of the the sequential pitch search, but the pitch searching time was reduced about 87 % as relative computation time.

REFERENCES

[1] A. N. Ince, *Digital Speech Processing*(speech coding, synthesis, and recognition), Kluwer Academic Publishers, 1992.

[2] W. B. Kleijn et al., "Fast Methods for the CELP Speech Coding Algorithm", IEEE Trans., Acoustics, Speech and Signal Processing, Vol.38, No.8, pp.1330 -1341, Aug. 1990.

[3] R. C. Rose and T. P. Barnwell, "Design an Performance of an Analysis-by-Synthsis Class of Predictive Speech Coders", IEEE Trans. Acoustics, Speech and Signal Processing, Vol.38, No.9, pp. 1489 - 1503, Sep. 1990.

[4] A. Le Guyader, D. Massaloux, and J. P. Petit, "Robust and Fast Code-Excited Linear Predictive Coding of Speech Signals", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1989.

[5] J. Menez, C. Galand, M. Rosso, and F. Bottau, "Adaptive Code Excited Linear Predictive Coder(ACELPC)", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1989.

[6] Joseph P. Campbell, Jr., Vanoy C. Welch, and Thomas E. Tremain, "An Expandable Error-Protected 4800 bps CELP Coder(U. S. Fedral Standard 4800 bps Voice Coder)", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1989.

[7] R. C. Rose and T. P. Barnwell III, "Quality Compression of Low Complexity 4800 bps Self Excited and Code Excited Vocoders", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1987.

[8] Grant Davidson and Allen Gersho, "Complexity Reduction Methods for Vector Excitation Coding", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1986.

[9] M. R. Schroeder and B. S. Atal, "Code-Exited Linear Prediction(CELP) : High-Quality at Low Bit Rates", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, pp. 25.1.1 - 25.1.4, 1985.

[10] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signal*, Prentice-Hall, 1978.

[11] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, Springer Verlag, New York, 1976.