Reduction of Pitch Searching Range in CELP vocoder

CELP 부호화기에서 피치검색범위의 단축

Jae Yeol Rheem*, Myung Jin Bae**, and Sou Guil Ann*

임 재 열*, 배 명 진**, 안 수 길*

ABSTRACT

The major drawback in Code Excited Linear Prediction(CELP) type speech coders is their large computational requirements. In this paper, a simple method which reduces the pitch searching time without degradation of quality in the correlation based pitch predictor is proposed. The basic idea is that, based upon the observational regularity of the correlation function in pitch search, the searching range can be restricted to the positive side by estimating the width of negative envelope with the width of previous positive envelope. By restricting the range of pitch search, required computations are reduced. Experimental result shows that about 40% reduction can be achieved by the proposed method without lowering the speech quality.

요 약

부호여기선형예측(CELP) 형태의 음성 부호화기에 있어서 주된 난점은 요구되는 계산량이 많다는 것이다. 본 논문에서는 음질의 저하없이 자기상관에 근거한 피치예측기에서 피치 찾는 시간을 줄이는 간단한 방법을 제안한다. 기본적인 발상은, 피치찾는 과정에서 나타나는 자기상관 함수의 특성에 근거하여, 음의 봉우리의 폭을 이전 양의 봉우리의 폭을 이용하여 추정하여 건너 뜀으로써 피치찾는 범위를 자기상관함수의 양의 값쪽으로 제한하는 것이다. 이렇게 피치찾는 구간에 제약을 가함으로써 요구되는 계산량은 감소하게 된다.

제안된 방법을 적용한 결과, 평균 약 40%의 계산량 감소를 음질의 저하없이 얻을 수 있었다.

I. INTRODUCTION

After the introduction of Code Excited Linear Prediction(CELP) speech coder in 1984 [1], many researches have been performed to achieve high quality speech below 4.8kbps within reduced computational requirements. The major drawback in CELP type analysis-by-synthesis speech coders is their large computational requirements in codebook and pitch searches [2]. CELP analysis consists of three basic functions : 1) short delay spectrum prediction, 2) long delay pitch search, and 3) residual codebook search. Spectrum analysis is performed once per frame by open-loop, usually 10th order autocorrelation LPC analysis using no preemphasis and 15 Hz bandwidth expansion with a Hamming window [3]. Codebook search is performed by closed-loop analysis using conventional minimum mean squared prediction error cri-

*Dept. of Electronics Engr., Seoul National University
**Dept. of Information and Telecomm. Engr., SoongSil University
접수일자 : 1993. 3. 30

terion of the perceptually weighted error signal. Pitch search is usually performed by closed-loop analysis using one of the following: filtering [4], self-excited [5], or adaptive codebook [6] methods. Since Pitch search is performed four times per frame based upon analysis by synthesis technique and all of the available pitch lags are exhaustively searched, it requires great computations. The computations of pitch search and codebook search are similar.

In this paper, a simple method which reduces the pitch searching time without degradation of quality in the correlation based pitch predictor is proposed. Based upon the observational regularity of the correlation function in pitch search, the searching range can be restricted to the positive side by estimating the width of negative envelope with the width of previous positive envelope. By restricting the range of pithch search, required computations are reduced. Experimental result shows that about 40% reduction can be achieved by the proposed methout lowering the speech quality.

Ⅱ. PITCH SEARCH IN PITCH FILTER

Fig. 1 shows a typical flow for pitch search using one-tap pitch filter. Pitch search is performed based analysis-by sysnthesis technique to select parameters such as the pitch lag L and pitch gain b for pitch prediction filter which minimize the error between the input speech and the synthesized speech. In Fig. 1, ZIR is zero input response and α is perceptual weighting constant and $A(z)/A(z/\alpha)$ is a perceptual weighting filter. Pitch synthesis filter is given as

$$\frac{1}{P(z)}=\frac{1}{1-bz^{-L}} \tag{1}$$

When $x(n)$ is the perceptually weighted input speech $y(n)$ is the perceptually weighted synthesized speech, the mean squared error(MSE) equation through pitch filter is

$$\begin{aligned} MSE &= \frac{1}{L_P}\sum_{n=0}^{L_P-1}(x(n)-y_L(n))^2 \\ &= \frac{1}{L_P}\sum_{n=0}^{L_P-1}(x(n)-by(n-L))^2 \end{aligned} \tag{2}$$

where L_P is the length of pitch analysis frame. The objective is to choose the L and b which minimize the MSE. This is equivalent to maximizing

$$E_L=\frac{(Exy)^2}{Eyy} \tag{3}$$

where

$$Exy=\sum_{n=0}^{L_P-1}x(n)\ y_L(n)$$

$$Eyy=\sum_{n=0}^{L_P-1}y_L(n)y_L(n).$$

The optimum b for the given L is found to be

$$b_L=\frac{Exy}{Eyy} \tag{4}$$

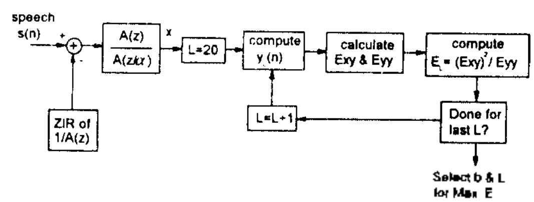

Fig 1. An example of implementation flow for pitch search

This search is repeated for all allowed values of L(usually from 17 to 143 or from 20 to 147). The lag L and the pitch gain b that maximize E_L are chosen for transmission.

Since ptich search is done four times per frame by this exhaustive search(every 5 or 7.5 msec), it requires very large computations. To reduce the burden, several methods such as recursive convolution [3][7] and approximations of correlation function [4][8] and Delta search [3] are used. Delta search method exploits the natural smoothness of pitch lag. For odd subframes, all of available lags searched while for even subframes, only

32 lags relative to the previous subframe are searched. Delta search greatly reduces computational complexity and data rate while causing no perceivable loss in speech quality.

However, the formulation used for the pitch filter is such that it removes long-term correlations, whether due to actual pitch excitation or not (strictly it is not a true pitch estimator), the chosen pitch lag L can be an improper lag even for voiced speech and shows doubling and halving (i.e. submultiples of pitch lag) frequently as shown in Fig.2.

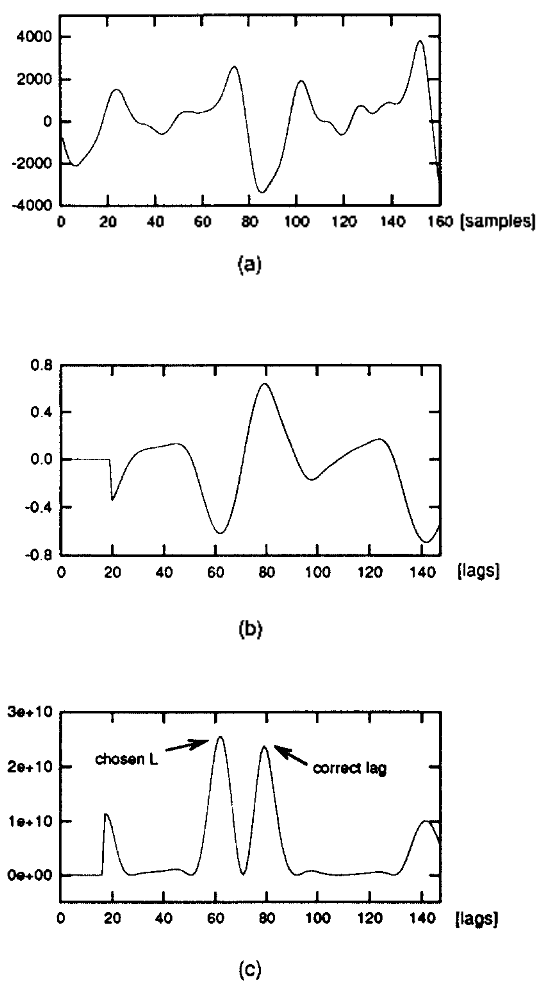

Fig 2. An example of correlation function.
(a) speech data (b) b = Exy/Eyy
(c) $E_L = (Exy)^2/Eyy$

To overcome this practically, the *MSE* criterion is usually modified to check the error at submultiples of the lag to determine if it is within an allowable level of *MSE* [3]. The optimum ***b*** can be restricted to be positive [7]. In that case, E_L which produces a negative is ignored in the search.

Ⅲ. PROPOSED PITCH SEARCH METHOD

In connection with correlation based pitch estimation method, the true pitch lag for voiced speech is always located at the peak of a positive envelope in the correlation function [9]. Based upon this fact, pitch lag search in prediction filter can be done on the correlation function and the search range can be restricted to the positive side of correlation function, if possible [7].

The correlation function shows some regularity and has the following properties. The envelope of correlation function changes slowly, for speech signal is highly correlated. The positive and negative envelopes are alternative and the width of each envelope is usually maintained by the effect of the first formant of voiced speech.

Based upon the properties of correlation function, the width of a negative envelope can be estimated by the width of the previous positive envelope. By skipping the lags corresponding that width, pitch search time reduction can be achieved.

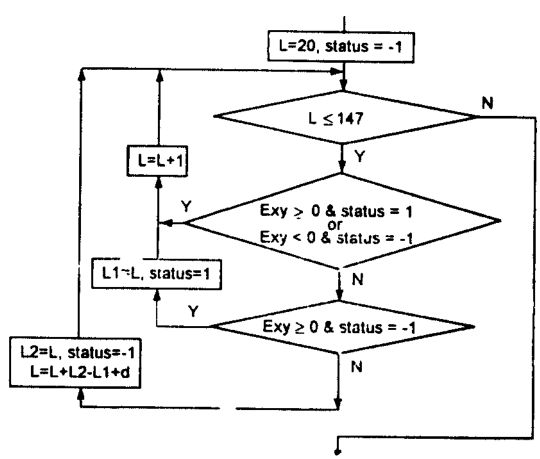

Fig 3. Proposed pitch search algorithm.

Since the positive peaks of correlation function are maintained, the performance does not change. Fig.3 shows the flow of the proposed algorithm, where $d(\geq 1)$ is adjusting constant for width skipping.

Ⅳ. EXPERIMENTAL RESULT

For experiment, four sentences pronounced five times by two males and one female speakers were used for test data base. The speech signal was sampled at 8kHz and lowpass filtered and digitized with a 16 bits A/D converter. We used a 20 ms frame size with four 5 ms subframes. Spectrum analysis was performed once frame by open-loop, 10th order autocorrelation LPC analysis using no preemphasis with a 20 ms Hamming window. In perceptual weighting, we chose $\gamma = 0.8$ and in pitch search, lags from 20 to 147 were searched and $d = 5$. In our experiment, real time implementation was not considered.

Fig.4 shows the correlation functions obtained by the conventional full search and the proposed method. The positive envelopes of the two correlation functions are the same. The selected pitch lag L and pitch gain b are not affected by the proposed method and the quality of speech is preserved. Since the shape of the correlation function varies frame to frame, the required amount of computations for pitch search varies. However, the proposed method works, for speech signal is highly correlated and almost half of the correlation function is negative envelopes.

To compare the pitch search methods, average search time was calculated in personal computer for test speech data base. As a result, average pitch search time, 12.2 sec was obtained by the conventional exhaustive search and 7.4 sec by the proposed method. Thus average 39.3% reduction of pitch search time was achieved for test speech database.

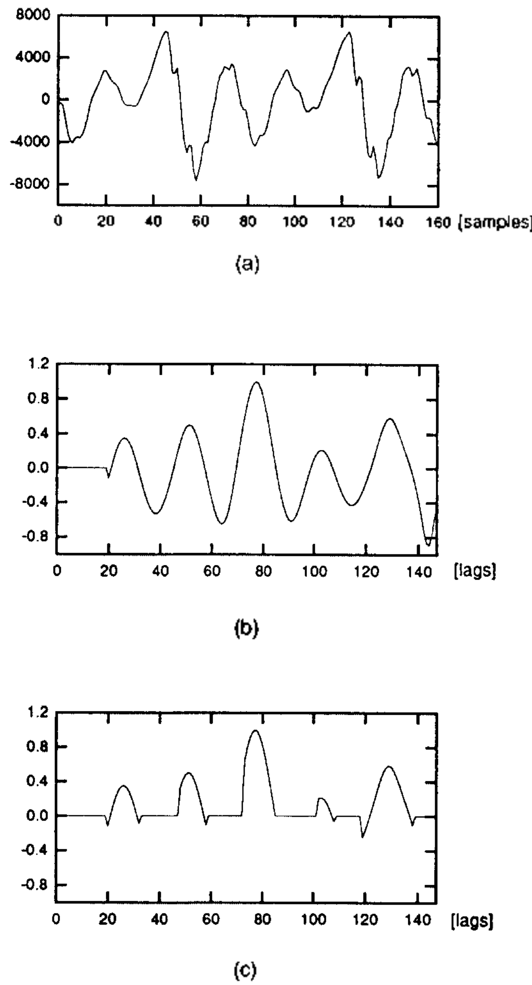

Fig 4. An example of the proposed method.
(a) speech data
(b) Exy/Eyy by the exhaustive search
(c) Exy/Eyy by the proposed method

Ⅴ. CONCLUSION

In this paper we proposed a simple pithod which preserves the quality of CELP vocoder with reduced complexity. The basic idea is to restrict the pitch search range to positive envelopes by estimating the width of negative envelope with that of previous positive envelope in the correlation function. Employing the proposed method, we can get approximately 40% complexity reduction in the pitch search.

Since the proposed method is performed in each subframe, great reduction of computational complexity can be expected by combining the proposed method with the Delta search method.

REFERENCE

1. B.S. Atal and M.R. Schroeder, "Stochastic Coding of Speech at Very Low Bit Rates," *Proceedings of ICC*, pp.1610-1613. 1984.
2. M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction(CELP) : High Quality at Low Bit Rates," *Proceedings of ICASSP85*, pp.937-940, 1985.
3. J.P. Campbell, Jr., V.C. Welch and T.E. Tremain, "An Expandable Error Protected 4800 bps CELP Coder(U.S. Federal Standard 4800 bps Voice Coder)," *Proceedings of ICASSP89*, pp.735-738, 1989.
4. R.P. Ramachandran and P. Kabal, "Pitch Prediction Filters in Speech Coding," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-37, no.4, pp.467-478, April 1989.
5. R. Rose and T. Barnwell, "Quality Comparson of Low Complexity 4800 bps Self Excited and Code Excited Vocoders," *Proceedings of ICASSP87*, pp. 1637-1640, 1987.
6. D. Lin, "Speech Coding Using Efficient Pseudo Stochastic Block Codes," *Proceedings of ICASSP87*, pp.1354-1357, 1987.
7. Vocoder software : high level design ver. 1.0, Qualcomm inc., Oct.1990.
8. J. Menez, C. Galand, M. Rosso, and F. Bottau, "Adaptive Excited Linear Predictive Coder(ACELPC)," *Proceedings of ICASSP89*, pp.132-135, 1989.
9. L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signal*, Prentice Hall, 1978.

▲Jae Yeol Rheem

Jae Yeol Rheem was born in Korea on January 15, 1964. He received the B.S. and M.S. degrees in electronics engineering from Seoul National University in 1986 and 1988, respectively. He is currently working towards the Ph.D. degree at the Department of Electronics Engineering, Seoul National University, Korea. His research interests include communication theory, speech signal processing, spectrum estimation and adaptive signal processing.

▲Myung Jin Bae : Vol.11, No.1E

▲Souguil Ann : Vol.11, No.1E