개선된 델타검색기법을 이용한 피치검색시간의 단축

이주헌, 방시열*, 배명진*, 안수길

서울대학교 전자공학과, *숭실대학교 정보통신학과

AN ALGORITHM TO REDUCE THE PITCH SEARCHING TIME USING MODIFIED DELTA SEARCH IN CELP VOCODER

JooHun Lee, Bang Siyeol*, MyungJin Bae*, SouGuil Ann

Department of Electronics Engineering, Seoul National University, Seoul 151-742, Korea.
*Department of Telecommunication Engineering, Soongsil University, Seoul 156-743, Korea.

ABSTRACT

The major drawback in the Code Excited Linear Prediction(CELP) type vocoders is their large computational requirements. In this paper, a simple method is proposed to reduce the pitch searching time in the pitch filter almost without degradation of quality. On the basis of the observational regularity of the correlation function of speech, only the limited numbers of pitch lags are considered to be an optimum pitch. This is done by skipping the negative envelope side of the correlation function and limiting the maximum number of lags to be considered preliminarily. By doing so, we can reduce the computational time of pitch searching more than 51% with negligible quality degradation. In addition to that, by combining that method with the conventional delta search technique, we can reduce the computational time requirements more than 60% without serious lowering the speech quality in segmental SNR measure compared to the conventional full search method.

I. INTRODUCTION

After the introduction of the Code Excited Linear Prediction(CELP) speech coder in 1984 [1], there have been many researches 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. The 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]. The codebook search is performed by closed-loop analysis using conventional minimum mean squared prediction error criterion of the perceptually weighted error signal. The pitch search is done usually using one of the followings: filtering [4], self-excited [5], or adaptive codebook [6] methods. Since the 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 computational complexity. These computational requirements of the pitch search are almost same as those of the codebook search and time reduction of the pitch search can reduce the overall computational requirements in CELP considerably.

In this paper, a simple method is proposed to reduce the pitch searching time in the correlation based pitch predictor with audible distortion of speech quality. On the basis of the observational regularity of the correlation function in pitch search, the searching range can be restricted to the positive envelope 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 34% reduction can be achieved by doing so. In addition to that, by limiting the maximum number of available lags in pitch searching by a constant number preliminarily, more reduction can be achieved up to 51% almost without lowering the speech quality in segmental SNR measure. Finally, by combining those technique with the conventional delta search technique, more reduction of computational time requirements can be achieved up to more than 60% with negligible quality degradation

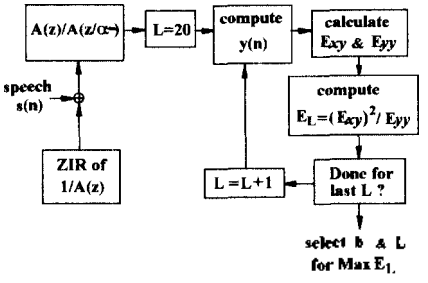

Fig 1. An example of implementation flow for pitch search.

II. DELTA SEARCHING IN PITCH FILTER

Fig. 1 shows a typical flow for pitch search using one-tap pitch filter. Pitch search is performed based on analysis-by-synthesis technique to select parameters such as the pitch lag L and pitch gain b for pitch prediction filter which minimize the weighted 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

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

x(n) and y(n) are the perceptually weighted input speech and the perceptually weighted synthesized speech, respectively. The mean squared error (MSE) equation through pitch filter is

$$\mathrm{MSE} = \frac{1}{L_p}\sum_{n=0}^{L_p-1}(x(n) - by_L(n))^2 \tag{2}$$

$$= \frac{1}{L_p}\sum_{n=0}^{L_p-1}(x(n) - by(n-L))^2$$

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

$$E_L = \frac{(E_{xy})^2}{E_{yy}} \tag{3}$$

where

$$E_{xy} = \sum_{n=0}^{L_p-1} x(n)y_L(n)$$

$$E_{yy} = \sum_{n=0}^{L_p-1} y_L(n)y_L(n).$$

The optimum b for the given L is found to be

$$b_L = \frac{E_{xy}}{E_{yy}} \tag{4}$$

For the conventional full search method, this search is repeated for all allowed values of L (usually from 20 to 147) by exhaustive manner. After that, the lag L and the pitch gain b that maximize EL are chosen for transmission. Since pitch 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], 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. The delta search greatly reduces the computational complexity and data rate while causing no perceivable loss in speech quality.

III. PROPOSED PITCH SEARCH METHOD

In connection with the pitch estimation method based on the correlation, 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 long delay prediction filter can be done in the correlation function and the search range can be restricted to the positive envelope side of correlation function, if possible[7]. The correlation function shows some regularity and has the following properties. The envelope of correlation function varies 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[9].

Based upon the properties of correlation function as mentioned above, the width of a negative envelope can be estimated by the width of the previous positive envelope. By skipping the lags corresponding to that width, pitch search range reduction can be achieved. Since the positive peaks of correlation function are maintained, the performance in segmental SNR does not change. To adjust the skipped range larger, which brings out more reduction of computational time requirements, we introduce the adjusting constant d(≥ 1), which is multiplied to the width of the previous positive envelope.

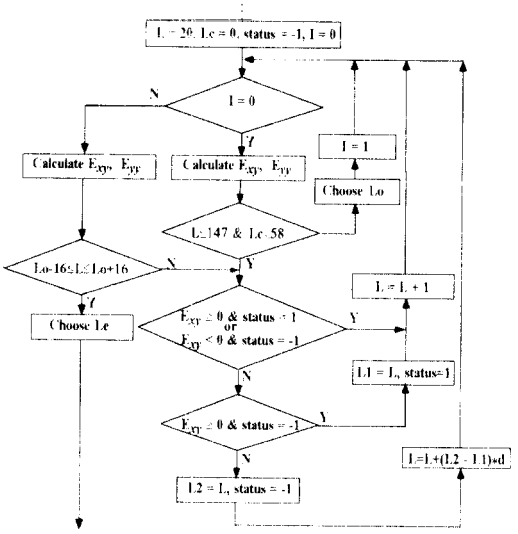

Fig. 2 Proposed pitch search algorithm.

Also, from the fact that, for voiced speech, both the numbers of positive correlation lags and negative ones are approximately same and that the maximum value of L is 147, the maximum number of available lags at which the correlation E_{xy} is calculated and checked can be limited. By counting the lags taken part in correlation computations with LC and restricting this maximum number by the threshold, L1, pitch searching time can be more considerably reduced with negligible degradation of speech quality in segmental SNR. This threshold is set preliminarily by a constant.

So, during the pitch search for one frame, if Lc get to L1, the rest of lags are not considered to be a proper L and the L search of that frame comes to end. In case of that the number of lags which have positive correlation is much more than 74 (=10+(128/2)), it is considered that the frame lies in the unvoiced or silence segment. Actually, when d is

set on 1.2, the maximum number of available lags can be reduced up to 58 (=128/(1+1.2)) almost without SNR degradation. As a result mentioned so far, our proposed method can reduce the overall pitch search time requirements remarkably though it introduces extra operations for Lc and d so to increase the time burden little.

In addition to the method described above, if we combine the conventional delta search technique, more reduction of computational time requirements can be achieved. That is, by considering 32 lags of the even subframe relative to the optimum L of the previous odd subframe, the number of lags to be considered in one frame can be reduced. When we set LT on 58, the number of the exempted lags in one frame minimally reaches up to 52 (2*(58-32)), which is 22.4% of total lags to be considered in one frame (52/(58*4)).

Fig. 2 shows the flow of the proposed pitch search algorithm. L is the lag index which varies from 20 to 147 and Lc is the counted number of lags where the correlation of E_{xx} is calculated. Status points out whether the positive envelope appeared or not during pitch search up to present lag. Lo and Le are the chosen optimum lags in odd subframe and even subframe respectively. Index I is the index to point out whether that subframe is the even subframe.

IV. EXPERIMENTAL RESULT

For experiment, phoneme balanced five Korean sentences pronounced five times by four male speakers and one female speaker were used for test data base. The speech signal was sampled at 8kHz and lowpass filtered at 4kHz and digitized with a 16 bits A/D converter. We used a 20 ms frame size with four 5 ms subframes. For spectrum analysis, 10th order autocorrelation LPC analysis using no preemphasis with a 20 ms Hamming window was performed on every frame by open-loop. In perceptual weighting, we choose α = 0.8 and ,in pitch search, lags from 20 to 147 were searched. Under the above conditions, the segmental SNR and mean of computation time reduction ratio between the conventional full search method and the proposed method with various d and LT were obtained. The average required computation time of the conventional method and the proposed one for test speech data were measured and compared on personal computer(IBM 486DX-II)

From the results shown as the table, for d = 1.2 with various LT, the proposed method shows good performances. The more the number of lags to be considered is decreasing, the more computational time reduction is achieved while the more the speech quality degrades. However, until LT gets to less than 58, the degradation of SNR is so negligible to be perceived.

By using the proposed method, we can reduce the computational time requirements up to 62.3% compared to the conventional full search method with 0.32dB of degradation.

V. CONCLUSION

In this paper, we proposed a simple method which preserves the quality of CELP vocoder with reduced complexity. The basic idea of the proposed method is to consider only the limited number of lags in pitch searching by restricting the pitch searching range and combining with the conventional delta search technique. At first, we introduce the method to restrict the pitch search range to positive side of envelope in the correlation function and also to limit the number of available lags to be searched by proper constant. By doing so, the required pitch searching time can be greatly reduced with negligible degradation of speech quality. Those can be achieved by using several characteristics of speech signal such that the envelope of correlation function of speech signal varies slowly and the positive and the negative envelopes alternatively appear with maintaining the width of the previous envelope in a sufficiently short interval due to the first formant of voiced speech. In addition to that, we combined the conventional delta search technique with that method to achieve more reduced computational time requirements. While the skipping and limiting is performed for all available values, from 20 to 147 in odd subframes, that is performed for only 32 lags relative to the optimum L of the previous odd subframes in the following even subframes. By introducing the delta technique, we can reduce the bit rate as well as the computational time.

Employing the proposed method, we can get more than 60% complexity reduction in the pitch search.

REFERENCES

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 4800bps CELP Coder(U.S. Federal Standard 4800bps Voice Coder)," Proceedings of ICASSP 89, pp. 735-738, 1989.
4. R. P. Ramachandran and P. Kabal, "Pitch Prediction Filter 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 Comparison 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, Qualcomm inc., 1992.
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.
10. J. H. Lee and et al., "A Fast Pitch Searching Algorithm using Correlation Characteristics in CELP Vocoder", Proceedings of MILCOM'94, pp. 699-702, 1994.

Table 1. Comparison result between conventional full search method and proposed method with various LT when d = 1.2

Sentence number	Conventional full search (dB)	Proposed method (dB)					
		LT = 74	LT = 64	LT = 58	LT= 50	LT = 40	LT = 30
S1	11.67	11.20	11.10	11.06	10.86	9.83	9.56
S2	12.41	12.13	12.03	11.92	11.90	11.88	9.30
S3	11.95	11.83	11.81	11.79	11.71	11.54	11.46
S4	12.03	11.91	11.87	11.81	11.74	11.60	11.58
S5	11.52	11.50	11.50	11.43	11.41	11.38	11.17
Mean of SNR(dB)	11.92	11.71	11.66	11.60	11.52	11.25	10.61
Mean of time reduction ratio(%)	-	50.8	54.9	62.3	66.5	69.2	74.2

LT = maximum threshold of LC (samples/frame)