CELP 보코더에서의 고속피치검색을 위한 개선된 Skipping 기법

이주헌*, 배명진**, 안수길*
*서울대학교 전자공학과 **숭실대학교 정보통신공학과

The Skipping Technique: A Simple and Fast Algorithm to Find the Pitch in CELP Vocoder

Joohun LEE, MyungJin BAE and Souguil ANN

I. Introduction

Many potential applications for low bit rate speech coding algorithms demand good speech quality at a reasonable cost. The code excited linear prediction coding (CELP) has been a good alternative for this purpose to provide good speech quality at intermediate bit rate (4.8-9.6 kbps) [1]. However, it demands so large computation time to obtain high speech quality. This makes it impossible to realize the real-time implementation on low-cost hardware (one or two general purpose digital signal processing device such as DSP-16) in many applications such as a satellite-based mobile communication system. To overcome this problem, there have been many researches on reducing the computation time under the view of codebook design [2]. However, there have been relatively few researches on the pitch search time reduction, though the computation time in the pitch search is comparable with that in the codebook search and both of them take most of the computation time in CELP analysis [3].

In this paper, we propose a fast pitch search algorithm using modified skipping technique, which maintains high speech quality. The idea of the proposed skipping technique is to skip over the negative correlation lags of the full search range before computing their correlation values in the pitch search, since the optimum pitch is assumed to be found among the positive correlation lags. As a result, the pitch search is performed on the much reduced number of lags over the traditional full search method. To improve the speech quality with maintaining this reduced computation amount, we introduce modified skipping technique by using the reduced form of the correlation function.

The modified skipping technique achieved over 35 % reduction in pitch search time without speech quality degradation compared with the traditional full search method.

II. Skipping Technique in CELP Vocoder

CELP vocoder is mainly divided into two parts: the analysis part and the synthesis part. CELP analysis part consists of three basic functions: 1) short delay spectrum prediction, 2) long delay pitch search and 3) random or stochastic codebook search. While the spectrum analysis is performed once per frame by open-loop, the codebook search and the pitch search are done 4 times per frame by closed-loop respectively. Therefore, both the codebook search and the pitch search take up the most of the computation time in CELP analysis.

Pitch search is performed based on analysis-by-synthesis technique. Pitch search is to select the optimum pitch lag L and the pitch gain b for pitch prediction filter under the criterion of minimizing the weighted mean squared error between the input speech and the synthesized speech. Pitch synthesis filter is given as

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

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

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

where L_p is the length of pitch analysis frame and $y_L(n)$ is the synthesized speech for the given L. The object is to choose L and b which minimize the 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

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

In the traditional full search method, the calculation of the correlation E_{xy} is repeated exhaustively for every allowed value of L which ranges from 20 to 147 on every subframe (1/4 frame). Then the lag L and the pitch gain b that maximize E_L are chosen for transmission.

In the pitch estimation using the correlation, the true pitch lag for voiced speech is always located at one of the positive peaks of the envelope in the correlation function [5]. Based on this fact, pitch search can be done in the correlation function and the search range can be restricted to the positive intervals of the correlation function, if possible. In QCELP, the optimum value of b is restricted to be positive, which means that the correlation E_{xy} is positive [3]. This fact allows us to reduce the computation time simply by excluding such lags that have the negative value of E_{xy} from the computation beforehand. The following properties observed in the correlation function make it possible in a simple way.

The properties of the correlation function of speech signal in a short interval are as follows: (1) The envelope varies slowly for voiced speech, since the speech signal is highly correlated. (2) The positive and the negative intervals appear alternately. (3) The width of each interval is almost invariant because of the effect of the first formant of voiced speech. These properties make it possible to estimate the width of a negative interval by the width of the previous positive interval.

From those facts mentioned so far, the computation time can be reduced without degradation by excluding the negative correlation lags in the pitch search and this can be done in a simple way. A simple and efficient way of exclusion of those lags is to estimate the width of each negative interval and skip over as that interval by the estimated width beforehand. This can be done simply by counting the lags in a positive interval of the correlation function and then skipping over as many lags as the counted number in the following negative interval. This procedure is repeated sequentially in a subframe. This causes the reduction of pitch search range considerably. However, the performance in segmental SNR is still maintained since the positive peaks of the correlation function are not lost.

To reduce the search range more, a scaling constant d is multiplied to the counted number of each positive interval [6]. An appropriate value of d makes the skipped range larger without losing the peaks of the positive intervals.

III. Skipping Technique with Reduced Form of Correlation Function

To obtained the sign of the correlation value with reduced computation, the reduced form of the correlation function is used. This reduced form of the correlaiton function is defined as follows;

$$R(L) = \sum_{n=-1}^{1} s(n)s(n-L) + \sum_{k=-1}^{1} s(k)s(k-L). \tag{5}$$

In this equation, n and k are the indexes for the highest peak and the deepest valley in the speech signal. This equation can be derived from the fact the correlation value of the highest peak appears in as same manner as that of the entire signal. This is true in case of the correlation value of the deepest valley. Therefore, the correlation function obtained by the entire signal has the same features as that by only the highest peak and the deepest valley. Not to be affected by the impulse noise, the previous and the next one samples of the highest peak and the deepest valley are considered together in the equation. With this reduced form of the correlation function, the correlation value can be obtained with much reduced times of multiplication and summation. For example, when the subframe size is 60, the number of multiplication is reduced to 1/10 times compared with that of the original correlation function.

To maintain high speech quality when the skipping technique is applied, the reduced form of the correlation function is used. Regardless of the large computation reduction, the positive correlation pitch lags are sometimes skipped over together with the negative correlation pitch lags when the skipping occures. This results in speech quality degradation. Not to skip over the positive region, when the correlation value is positive after the skipping occures, the pitch search is performed backward until the negative correlation value appears. Since the decision of the sign of the correlation value is performed by the reduced form the correlation, this procedure needs little more computation. However, since the correlation values in the negative region are calculated by this reduced form of the correlation function, the entire computation load is reduced

IV. Result and Discussion

Five phoneme balanced Korean sentences are used as test data Each sentence is pronounced 5 times by three male and two female speakers. The speech signal is low-pass filtered at 3.4 kHz, sampled at 8 kHz and digitized with a 16 bits A/D converter. The frame size is 20 msec and each frame is divided into 4 subframes. For spectrum analysis, the 10-th order autocorrelation LPC analysis is performed on every frame. The perceptual weighting constant is 0.8.

The average required computation amounts of the skipping technique and the modified one for the test data are calculated by the equation in [4] and compared with that of the traditional full search method. As an objective quality measure, the average segmental SNR (SEGSNR) is computed.

For the test data, the average number of occuring such errors that the signs of both the original correlation function and the reduced form of the correlation function are different is 6 However, those errors occur only around the pitch lags where the correlation value changes from positive to negative or vice versa. Since their match scores are small, the optimum pitch can be exactly found without affected those errors.

For the skipping technique with the various values of d, the amounts of the reduced computation and their values of SEGSNR are represented in Table I. Also, the same experimental results for the skipping technique with the reduced form of the correlation function are presented in Table II.

As shown in both tables, the modified skipping technique can more improve the speech quality with the reduced computation compared with the simple skipping technique. Additionally, the larger value of d is selected, the more reduced computation time can be achieved. This results from the fact that the negative region can be skipped more perfectly when the value of d is larger.

V. Conclusions

From the above results, we conclude that the modification skipping technique can preserve high speech quality with reduced computation time over the traditional full search method.

The simple skipping technique may skip over the true pitch lag during the skipping procedure. However, the modified skipping technique does not miss the optimum pitch lag since the pitch search is performed for every potive correlation lag. Moreover, this can be done without increasing the computation load by using the reduced form of the correlation function.

Since the primary research goal is to find and develop a "near toll quality" speech coder to be used in a nationwide, satellite-based mobile communication system, the technique of interest is to maintain high quality at 4.8 kbps and operate acceptably in such a mobile communication environment which could be realized by using reasonable cost resources with low power dissipation. The experimental results show that the proposed algorithm is a good alternative for this purpose. The proposed algorithm is expected to be applied to any type of the conventional CELP coder without additional problems since it is simply derived from the conventional full search method.

References

[1] B. S. Atal and M. R. Schroder, "Stochastic Coding of Speech at Very Low Bit Rates", *Proceedings of ICC*, pp.1610-1613, 1984.

[2] W. B. Kleijn, D. J. Krasinski and R. H. Ketchum, "Fast Methods for the CELP Speech Coding Algorithm", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 38, no. 8, pp. 1330-1342, 1990.

[3] *Proposed EIA/TIA Wideband Spread Spectrum Standard: High Level Design*, Qualcomm inc., 1993.

[4] J. P. Cambell, V. C. Welch and T. E. Tremain, "The New 4800 bps Voice Coding Standard," *Proc.Military and Government Speech Tech'89*, 1989.

[5] L. R. Rabiner and R. W. Schafer, "Digital Processing of Speech Signal", Prentice Hall, 1978.

[6] J. H. Lee, M. J. Bae and S. G. Ann "A Fast Pitch Searching Algorithm Using Correlation Characteristics in CELP Vocoder", *IEEE Comm. Soc., Proc. MILCOM'94*, pp. 699-702, 1994.

Table I. SEGSNR and computaion amount results for the conventional full search method and the skipping technique with varying the value of d.

	Full Search	Delta Search	$d = 0.8$	$d = 1.0$	$d = 1.2$	$d = 1.5$
SEGSNR (dB)	11.92	11.47	11.90	11.87	11.81	11.25
Computation amount (MIPS)	3.5	2.3	2.49	2.40	2.30	2.18

Table Ⅱ. SEGSNR and computaion amount results for the conventional full search method and the skipping technique with varying the value of d.

	Full Search	Delta Search	$d = 0.8$	$d = 1.0$	$d = 1.2$	$d = 1.5$
SEGSNR (dB)	11.92	11.47	11.92	11.92	11.92	11.92
Computation amount (MIPS)	3.5	2.3	2.25	2.20	2.12	2.02