A Simple and Fast Pitch Search Algorithm Using a Modified Skipping Technique in CELP Vocoder

개선된 Skipping 기법을 이용한 CELP 보코더에서의 고속피치검색 알고리듬

Joo Hun Lee*, Myung Jin Bae*, Choon Woo Kwon**

이 주 헌*, 배 명 진*, 권 춘 우**

ABSTRACT

Based on the characteristics of the correlation function of speech signal, the skipping technique can reduced the computation time considerably with a little degradation of speech quality. To improve the speech quality of the skipping technique, we use the reduced form of the correlation function to check the sign of the correlation value before the match score is calculated. The experimental results show that this modified skipping technique can reduce the computation time in pitch search over 35% compared with the traditional full search method without quality degradation.

요 약

CELP 보코더에서의 피치검색에 있어서, 음성신호의 상관함수의 몇가지 특성을 이용한 스킵기법은 음질의 저하를 거의 일으키지 않으면서 피치검색에 필요한 계산량을 줄여 준다. 하지만, 때때로 양의 상관값을 갖는 검색대상지연들이 검색에서 제외되어 음질이 저하되는 것을 완전히 막을 수는 없다. 이러한 단점을 보완하기 위해서 약식의 상관함수식을 이용한 개선된 스킵기법을 제안한다. 제안된 방법은 기존의 전체검색방법보다 계산량을 35% 이상 줄이면서도 음질의 저하는 일어나지 않는다.

I. Introduction

Many potential applications for low bit rate speech coding algorithm demand good speech quality at a reasonable cost. The code excited linear prediction coding (CELP) has been a good alternative to provide good speech quality at intermediate bit rate (4.8-9.6 kbps) [1]. However, due to its large computation complexity for high speech quality, it is 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, some researches have been proposed to reduce the computation complexity in pitch search since pitch search needs large computation complexity along with the stochastic codebook search in CELP analysis [3]-[6].

In [7], the skipping technique is a good method to reduce computation complexity considerably without degradation of speech quality. The idea of the skipping technique is to skip over the negative correlation lags in the allowed full search range based on the characteristics of correlation function of speech signal. As shown in [7], with appropriate modifications, the skipping technique reduced the computation complexity considerably compared with the full search method. However, in this simple skipping technique, a little quality degradation of 0.3 dB is inevitable though it is not serious to be recognized.

In this paper, we proposed a modified skipping technique which maintains higher speech quality than

*Department of Telecommunication Engineering, Soongsil University

**Dong Yang Technical College Computer & Office Automation

접수일자 : 1995년 3월 10일

the simple skipping technique. By using the reduced form of the correlation function proposed in this paper, the modified skipping technique can avoid the whole possibility to skip the positive correlation lags in the simple skipping technique. The modified skipping technique achieved over 35% reduction in pitch search time without speech quality degradation compared with the traditional full search method.

Ⅱ. Pitch Search 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.

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

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

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

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

where L_p is the length of a subframe and $y_L(n)$ is the synthesized speech for the given L. The objective is to choose L and b which minimize the MSE and this is equivalent to maximizing

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

where $E_{xy}=\sum_{n=0}^{L_p-1} x(n)\,y_L(n), \quad E_{yy}=\sum_{n=0}^{L_p-1} y_L(n)\,y_L(n).$

The optimum b for the given L is found to be

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

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

Ⅲ. The Skipping Technique

According to the fact that the true pitch lag for voiced speech is always located at one of the positive peaks of the envelope in the correlation function [8], it is preferable for pitch search to be done only on the restricted numbers of lags the correlation values of which are positive among all allowed lags, if possible. In QCELP, the optimum value of b is restricted to be positive, which means that the correlation E_{xy} is positive [2]. This fact allows us to exclude such lags that may have the negative values of E_{xy} from pitch search. The following properties observed in the correlation function of speech signal make it possible in a simple way : (1) The envelope varies slowly for voiced speech, since the speech signal is highly correlated in a short interval. (2) The regions of positively correlated lags and the negatively correlated lags appear alternately. (3) The width of each region is almost invariant because the effect of the first formant of voiced speech is dominant. Based on the above properties, the width of each negatively correlated region can be estimated easilty by counting the number of lags in each previous positively correlated region. At the beginning of each negatively correlated region, by skipping over the following lags as many as the counted number, pitch search can be done only on the positively correlated lags without losing the positive peaks of the correlation function. This reduces considerably the computation complexity in pitch search without the degradation of speech number of each positive interval [7]. An appropriate value of d makes the skipped range larger without losing the peaks of the positive intervals.

Ⅳ. 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 correlation function is defined as follows :

$$R(l) = \sum s(n)s(n-l) + \sum s(k)s(k-l), \quad (5)$$

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 frame size is 60, the number of required 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 occurs. This results in speech quality degradation. Not to skip over the positive region, when the correlation value is positive after the skipping occurs, the pitch search is performed backward until the negative correlation value appears. Since the decision often sign of the correlation value is performed by the reduced form of the correlation, this procedure needs a 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.

V. Experimental Results and Discussions

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 time of the proposed method for the test data is measured 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 occurring such errors that the signs of both the original correlation 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 te-

Table I. SEGSNR and computation amount results for the conventional full search method and the simple 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 computation amount results for the conventional full search method and the modified 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

chnique can more improve the speech quality with the reduced computation compared with the simple skipping technique. Additionally, the large value 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.

Ⅵ. Conclusions

From our experimental results, we conclude that the modified 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 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 positive 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 problem 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. *Proposed EIA/TIA Wideband Spread Spectrum Standard High Level Design*, Qualcomm inc., 1993.
3. J. P. Cambell, V. C. Welch and T. E. Tremain, "An Expanded Error Protected 4800 bps CELP Coder (U.S. Federal Standard 4800 bps Voice Coder)," *IEEE Comm Soc., Proc. ICASSP'87*, pp. 735-738, 1989.
4. R. Salami, C. Laflamme and J. P. Adoul, "8kbit/s ACELP Coding of Speech with 10ms Speech-frame : A Candidate for CCITT Standardization," *Proc. ICASSP '94*, II-97, 1994.
5. A. Kataoka, T. Moriya and S. Hayashi, "Implementation and Performance of an 8kbit/s Conjugate Structure CELP Speech Coder," Proc. ICASSP'94, II-94, 1994.
6. K. Ozawa, M. Serizawa, T. Miyano, T. Nomura, M. Ikekawa and S. Taumi, "M-LCELP Speech Coding at 4kb/s with Multi-Mode and Multi-Codebook," IEICE Trans. Commun, vol. E77-b, no. 9, pp. 1114-1121, 1994.
7. 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.
8. L. R. Rabiner and R. W. Schafer, "Digital Processing of Speech Signal," Prentice Hall, 1978.

▲이 주 헌(Joo Hun Lee)
현재 : 숭실대학교 생산기술연구소 연구원
13권 2E호 참고

▲배 명 진(Myung Jin Bae)
현재 : 숭실대학교 정보통신공학과 교수
13권 2E호 참고

▲권 춘 우(Choon Woo Kwon)
현재 : 동양공업전문대학교 교수