A Pitch Detection Algorithm for Ultra-low-bit-rate Vocoder

Li Ye, Jiang Jingsai, Hao Qiuyun, Fan Yanhong,
Shandong Computer Science Center
Shandong Provincial Key Laboratory of computer Network
Jinan, China
liye@keylab.net

Ma Xiaofeng, Guo Qiang
Shandong Computer Science Center
Shandong Provincial Key Laboratory of computer Network
Jinan, China

***Abstract*—Pitch is one of the most important parameters to describe the speech characteristics. The paper proposed a new pitch detection algorithm for ultra-low-bit-rate vocoder based on dynamic programming as well as frequency-time domain analysis to eliminate the multi-half pitch errors. The algorithm detected the candidate pitches based on auto correlation function firstly and then utilized the dynamic programming and frequency-time domain analysis to get the final pitch. Simulation result showed that the algorithm had a better performance and could eliminate the multi-half pitch errors efficiently.**

Keywords-Pitch detection, ultra-low-bit-rate, dynamic programming, frequency-time domain analysi

I. INTRODUCTION

Speech communication is among the most important interfaces between human. In recent years, the speech coding rate become lower and lower. Speech coding algorithms with rate below 2.4kbps, which are also called ultra-low-bit-rate speech coding algorithms, become one of the most important research topics[1-6]. And many speech coding models were proposed, such as mixed excitation linear prediction (MELP) model, multi-band excitation (MBE) model, sinusoidal transform coding (STC) and waveform interpolation (WI). As Figure1 to Figure4 show, in these algorithms, pitch is among the most important parameters in the final quality of the synthetic speech.

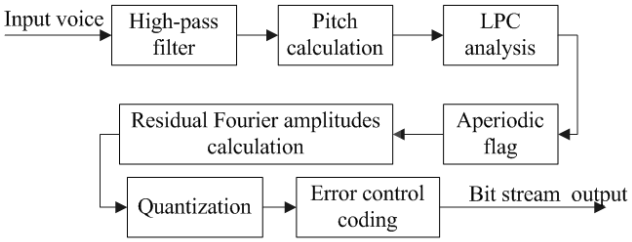

Figure 1. MELP Coding Algorithm

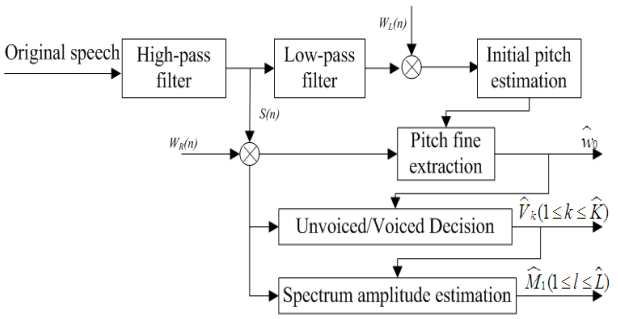

Figure 2. IMBE Coding Algorithm

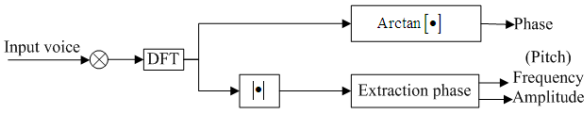

Figure 3. STC Coding Algorithm

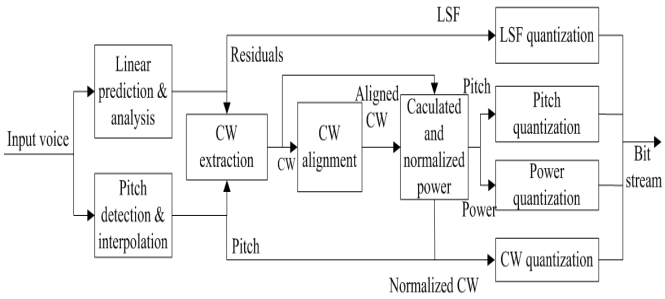

Figure 4. WI Coding Algorithm

Many pitch extraction algorithms have been proposed because of its importance [7-12]. However, these algorithms have some limitations because of some difficulties: the harmonic structure of speech signal varies widely; the head as well as end of a voiced period is difficult to judge; the pitch could change seriously some time itself. To overcome these problems, the paper proposes a new algorithm based on dynamic programming as well as frequency-time domain analysis. Experiments show that method could greatly improves the hearing quality of synthetic speech .

Published by Atlantis Press, Paris, France.

II. THE NEW PITCH DETECTION ALGORITHM

The bit number for pitch quantization in ultra-low-bit-rate speech coding algorithms is limited, so the integer pitch is detected in this algorithm instead of fraction pitch. So the computation cost would be reduced while the speech quality wouldn't be influenced too much.

In the algorithm, the frame length is 20ms and the sample rate is 8 KHz. The speech frame firstly passes through a low-ass filter whose cut frequency is 800Hz to remove the influence on pitch detection by high frequency signal. Because of the interacts between the glottis and track, the formant would affect the pitch detection. It's required to smooth the speech spectrum before pitch detection. In order to do this, the algorithm then imposes a two order linear predictive inverse filter on the speech signal. The final signal we got was referred to as S_j.

The algorithm utilizes the auto correlation function of s_j to detect the pitch parameter. Frequency-time domain analysis as well as the dynamic programming are then used to eliminate the multi-half pitch error.

A. Frequency-time domain analysis

Suppose that the auto correlation function of s_j is $r(i)$, then the peak value $p(i)$ would be detected firstly, which satisfies the condition as follows:

$$p(i)=\begin{cases} r(i), s.t.\ \ r(i)>r(i-1)\ \&\&\ r(i)>r(i+1) \\ 0, else \end{cases} \tag{1}$$

The array $p(i)$ would be used to detect the candidate pitches. To avoid the multi-half pitch error, some correction would be done firstly.

For each index i that Satisfies

$$r(i)>r(i-1)\ \&\&\ r(i)>r(i+1) \tag{2}$$

The value $p(j)$ would be revised. The index j should be a value between 20 and 40 which could be obtained by divide the index i by k, where k belongs to the close zone [2, 8]. The revise should be as follows:

$$p(j)=r(i)-0.00001\text{f},\ \ if\ p(j)<r(i)\ \&\ r(j)>thr\cdot r(i) \tag{3}$$

We get the value thr through steps as follows:

$$thr=\begin{cases} 0.5, & if\ r(i)>1.0\ \&\ i\geq 70 \\ 0.55, if\ r(i)>1.0\ \&\ i<70 \\ 0.95, if\ r(i)>0.8\ \&\ i\geq 70 \\ 0.97, if\ r(i)>0.8\ \&\ i<70 \\ 1.0, & else \end{cases} \tag{4}$$

After that, the frequency power analysis is done if the index i is larger than 40. The algorithm would get the frequency power E_1 and E_2 at both the index i and the index $i/2$. E_1 and E_2 could be obtained according to:

$$E_1=(\sum_{j=0}^{160-1}s_j*\cos(2\pi*j/i))^2+(\sum_{j=0}^{160-1}s_j*\sin(2\pi*j/i))^2$$
$$E_2=(\sum_{j=0}^{160-1}s_j*\cos(4\pi*j/i))^2+(\sum_{j=0}^{160-1}s_j*\sin(4\pi*j/i))^2 \tag{5}$$s

The revise of $p(i)$ should be as follows:

$$p(i)=\begin{cases} 0, if\ \begin{cases} E_1<0.01*E_2 \\ r(i)>0.6 \\ u_l=0\ or\ |i-p_l|>5 \end{cases} \\ p(i) \end{cases} \tag{6}$$

Where, p_l is the pitch of the last frame and u_l is the unvoiced/voiced flag of last frame. u_l is determined as follows:

$$u_l=\begin{cases} 0, \text{unvoiced} \\ 1, \text{voiced} \end{cases} \tag{7}$$

B. Dynamic Programming

The algorithm utilizes the dynamic programming to smooth the pitch parameter. The five max values in array P would be reserved, as Figure 5 shows.

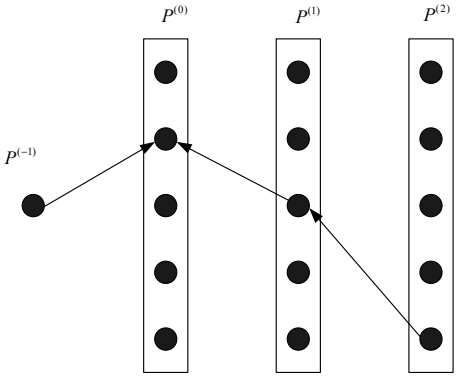

Figure 5. Dynamic Programming of Pitch

In order to get the pitch of the current frame, two future frames as well as one last frame would be considered. The candidate pitch of the current frame and future frame would be less than five, supposed as $R^{(k)}(n)$, where $k\in\{-1,0,1,2\}$ and $n\in\{0,1,2,3,4\}$. The corresponding value of $p(i)$ for candidate pitch is marked as $R^{(k)}(n)$。The pitch of the last frame is $P^{(-1)}$, the corresponding value of

array P is $R^{(-1)}$。The cost function for each pitch $P^{(l)}(j)$ is：

$$C_n^{(l)}(j) = c \cdot (1 - R^{(l)}(j)) \quad (8)$$

Where, c is the weight coefficient. The cost function between the adjacent pitch $P^{(l)}(i)$ and $P^{(l+1)}(j)$ is：

$$C_p^{(l)}(i,j) = \left| P^{(l)}(i) - P^{(l+1)}(j) \right| / \max(P^{(l)}(i), P^{(l+1)}(j)) \quad (9)$$

The dynamic programming aims at finding a pitch path that has the least cost among all possible paths. The pitch of the current frame in this path is then determined as the real pitch. The details of this dynamic programming could be referred to as reference [6].

III. SIMULATION RESULTS

The new algorithm is tested by a database including male as well as female speech. The algorithm LI proposed in his thesis for master degree was compared to because it is also designed for ultra-low-bit-rate speech coding algorithms [12]. We compare our algorithm with the algorithm proposed by LI.

The speech coding rate of the test algorithm is 600bps based on MELP. In the algorithm, the analysis frame length is 20ms and 6 frames are packed into one super-frame. The parameter quantization is based on the super-frame, so the bit number to quantize the all parameters is 72 . The bit allocation was shown in Table 1.

Table 1. Bit allocation for 600bps coding algorithm

Parameter	Bit Num
Band pass voicing strength	4
Pitch Parameter	11
Linear spectral parameters	47
Energy parameter	10
Residual Fourier magnitudes	0
Total	72

In the test, it could be seen that some pitch detection errors which occurs in his algorithm were corrected in our algorithm. For example, a pitch-double error in the old algorithm was corrected in the new algorithm. The original signal is “GUO LI DE JI YU” spoken by a male in Chinese. Clearly, the old algorithm brings a pitch detection error in front of the signal while the new algorithm eliminates it, as the Fig 6 to Fig 8 showed. This could improve the hearing quality of synthetic speech with subjective listening greatly.

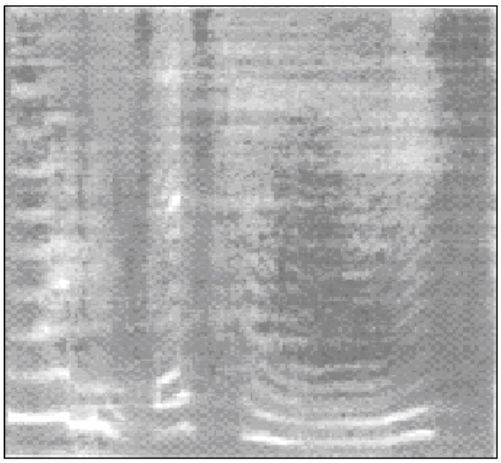

Figure 6. The spectrum of the original speech

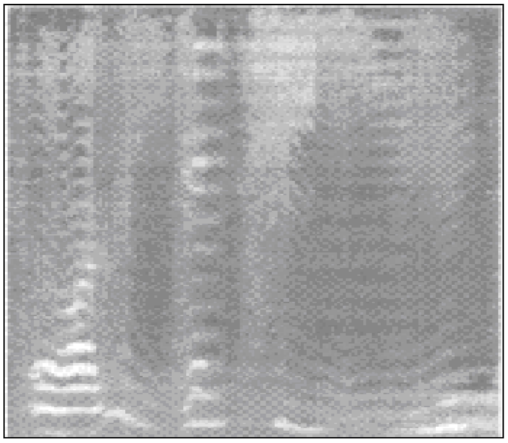

Figure 7. The spectrum of the old synthetic speech

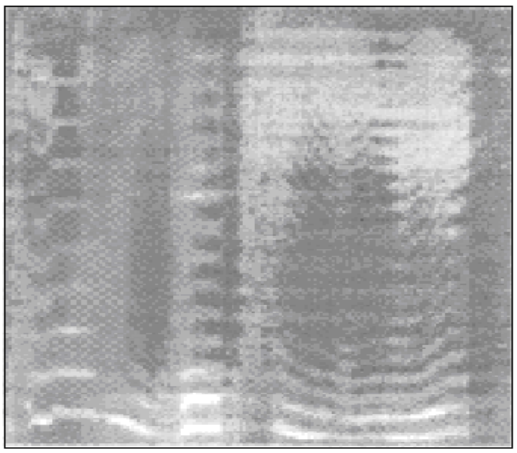

Figure 8. The spectrum of the new synthetic speech

We also test the mean opinion score (MOS) of the synthetic speech produced by both 600bps algorithms using the International Telecommunication Union (ITU) tool P.862. The MOS of the old algorithm is 2.579 and ours is 2.582.

Although there is only a little improvement, the place where the pitch was corrected sounds more naturally and clearly.

IV. CONCLUSION

Pitch is among the most important parameters for speech coding algorithms. The paper proposed a new pitch detection algorithm for ultra-low-bit-rate speech coding algorithms. Both dynamic programming and time-frequency domain analysis were used. Compared to the old algorithm, some pitch detection error could be eliminated clearly and the quality of synthetic speech could be improved much.

Experiments show that method proposed here could increase the MOS of the reconstructed speech of one 600bps vocoder based on MELP by 0.003 and greatly improves the hearing quality of synthetic speech with subjective listening. At present, the new pitch extraction algorithm has been applied to 300-2400bps vocoders based on MELP.

ACKNOWLEDGMENT

This work is supported by Natural Sciences Foundation of China (Grant No. 60802030).

REFERENCES

[1] Gottesman O., Gersho A.. Enhanced Waveform Interpolative Coding at Low Bit-Rate. IEEE Transactions on Speech and Audio Processing, 2001, 9(8):786~798.

[2] Tian Wang, Kazuhito Koishida, “A 1200/2400 BPS CODING SUITE BASED ON MELP”, Speech Coding IEEE Workshop Proceedings, 6-9 Oct 2002, pp: 90 – 92

[3] Tian Wang, Kazuhito Koishida, Vladiinir Cuperman, Allen Gersho, “A 1200 BPS SPEECH CODER BASED ON MELP”, IEEE International Conference on Acoustics, Speech, and Signal Processing, June 2000, Volume 3, 5-9, pp: 1375 – 1378

[4] McCree A, Barnwell T. P. III. “A mixed excitation LPC vocoder model for low bit rate speech coding”, IEEE trans. Speech Audio Process., 1995, 3(4), pp. 242-250

[5] McCree A, Truong K, George E B, et al. A 2.4kbit/s MELP Coder Candidate for the New US Federal Standard[C]. In: Proceedings of IEEE ICASSP, Atlanta, U S,1996. 200~203

[6] ZHAO Ming. Research on ultra low bit rate speech coding techniques and algorithms[D]. Tsinghua University, Beijing, 2004. (in Chinese)

[7] Rabiner L, Cheng M. A comparative performance study of several pitch detection algorithms[J] . IEEE Trans On Acoustics, Speech, and Signal Processing, 1976, 24(5) :399 - 418.

[8] Secrest B, Doddington G. Post processing techniques for voice pitch trackers [C]// International Conf On Acoustics, Speech, and Signal Processing. Paris: IEEE, 1982: 172 -175.

[9] Ney H. A dynamic programming technique for nonlinear smoothing[C]// International Conf On Acoustics, Speech, and Signal Processing. Atlanta: IEEE, 1981: 62 - 65.

[10] Ahmadi S, Spanias A S. Cepstrum-based pitch detection using a new statistical V/UV classification algorithm. IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 3, pp. 333 –338, May 1999.

[11] LIU Jian, ZHENG Fang. Combined magnitude difference function based pitch tracking algorithm [J]. J Tsinghua Univ (Sci & Tech) , 2006, 46(1) : 74 - 77.

[12] LI Junlin. Research on low bit rate speech coding algorithm:[Thesis for master degree]. Peking: Tsinghua University, 2004.