2002 International Computer Symposium (ICS2002)

Workshop on Artificial Intelligence

paper title :

Weighted Autocorrelation Function for Pitch Extraction from Noisy Speech

Wei-Wen Hung
洪偉文
Department of Electrical Engineering,
Ming Chi Institute of Technology,
明志技術學院電機系
84 Gungjuan Road, Taishan, Taipei,
Taiwan 24306, Republic of China.
臺北縣泰山鄉貴子村工專路八十四號
E-mail : wwhung@ccsun.mit.edu.tw
FAX : 886-02-2906-1780
Tel. : 886-02-2906-0379

Weighted Autocorrelation Function for Pitch Extraction from Noisy Speech

Wei-Wen. Hung

Abstract

In this paper, we discuss the use of a fuzzy weighting scheme to improve the accuracy of pitch extraction from noisy speech. The proposed fuzzy weighting scheme emphasizes the peak structure of the autocorrelation function (ACF) calculated from a speech frame while attenuating the components of ACF with smaller values. Experimental results show that by properly adjusting the fuzzy factor, the ACF-based fuzzy weighting scheme can provide higher capability in extracting the pitch information than some cepstrum-based and autocorrelation-based approaches in a noisy environment.

Key words : Pitch extraction, Autocorrelation function, Fuzzy weighting scheme.

1. Introduction

In a speech analysis-synthesis system, it is essentially necessary to extract pitch period in parallel with some spectral parameters. For this reason, many approaches to extract the pitch period of speech signals have been proposed [1]. Recent pitch extraction research has been undertaken from three viewpoints. One is how to reliably extract the periodicity of quasi-periodic signals. Another is how to correct the pitch extraction error owing to the disturbance of periodicity. The other is how to remove the vocal tract effects. Major errors in pitch extraction include double-pitch and half-pitch errors. Generally speaking, the pitch extraction (PE) methods can be roughly classified into three categories, i.e., waveform domain, spectral domain and correlation domain [2]. In the waveform domain, various logical processing is employed to remove superfluous waveform data and leave only pitch pulses. As to the spectral domain, pitch period is detected by tracking the peaks embedded in the spectral envelope. Among the three categories, autocorrelation-based method is known to be comparatively robust against background noises [3]. It utilizes autocorrelation function (ACF) of a waveform for periodicity detection and to be widely used in digital signal processing of speech. In this paper, a fuzzy weighting scheme (FWS) is incorporated into the calculation of an ACF. By means of emphasizing the peak structure of an ACF, it is expected that the accuracy of pitch extraction based on the ACF can be improved.

The remainder of this paper is organized as follows. Section 2 describes the formulation of the proposed fuzzy weighting scheme for pitch extraction from a noisy environment. In Section 3, we first introduce the database we used for conducting a series of simulation experiments. After that, we confirm the effectiveness of our method by comparing with some conventional methods based on several experimental results. Finally, we conclude the proposed method.

2. A Fuzzy Weighting Scheme for Pitch Extraction

The autocorrelation function $r(^2)$ is generally calculated by

$$r(\tau)=\frac{1}{N}\cdot\sum_{n=0}^{N-1}S(n)\cdot S(n+\tau)\,,\tag{1}$$

where $S(n)$ represents the analysis frame of a speech signal with N speech samples and $0\le\tau\le N-1$. The characteristic of $r(\tau)$ is that it attains its maximum value at $\tau=0$. If $S(n)$ has a period of T, then $r(\tau)$ has peaks at $\tau=n\cdot T$, where n is an integer. Essentially, $r(\tau=0)$ gives the largest value among $r(\tau=n\cdot T),\ \ for\ \ n=0,1,2,\cdots,(N-1)$. Other peaks of $r(\tau)$ usually decrease as τ increases. When the speech segment $S(n)$ is contaminated by background noises, it is possible that some abnormal peaks may occur and result in a half-pitch or double-pitch error. For such reasons, it is necessary to emphasize the true peaks of ACF by properly weighting the components in $r(\tau)$. Tetsuya and Hajime [4] validated the fact that the additive noise included in the ACF behaves independently with that included in the average magnitude difference function (AMDF), and to weight the ACF by the reciprocal of AMDF. Thus, the weighted ACF $\hat{r}(\tau)$ is expressed as

$$\hat{r}(\tau)=\frac{1.0}{[\text{¤}(\tau)+1.0]}\cdot r(\tau)\,,\tag{2}$$

where $\text{¤}(\tau)$ denotes the AMDF of the speech segment $S(n)$ and can be formulated as

$$\text{¤}(\tau)=\frac{1}{N}\cdot\sum_{n=0}^{N-1}\left|S(n)-S(n+\tau)\right|.\tag{3}$$

In this paper, we investigate the effect of weighting the ACF by means of a fuzzy approach. The proposed fuzzy weighting scheme (FWS) assigns to each component of the ACF a membership value between zero and one that indicates to what extent a particular component is emphasized. The membership function associated with the FWS can be of the form

$$\hat{r}(\tau)=\frac{\tau}{\sum_{\text{©}=0}^{N-1}\text{©}}\cdot r(\tau)\quad\text{and}\quad \text{©}=\sum_{\sim=0}^{N-1}\left(\frac{[\text{¤}(\text{©})+1.0]^{-1}}{[\text{¤}(\sim)+1.0]^{-1}}\right)^{\frac{1}{F-1}}.\tag{4}$$

In above equations, the “fuzziness” of the proposed weighting scheme is controlled by the fuzzy factor F which is greater than unity. By properly adjusting the fuzzy factor, we can achieve various extents of fuzziness for the FWS. When the fuzzy factor tends to unity, and $r(i)$ has the largest value among $r(j),\ for\ \ 0 \le j \le N-1$, then the weights associated with those ACF components are distributed with $w(i)=1.0$, and $w(j)=0.0\ \ for\ \ 0 \le j \le N-1,\ \ and\ \ j \ne i$. On the other hand, in the case of $F \to \infty$, all the weights become equal, that is $w(i)=1.0/N,\ \ for\ \ 0 \le i \le N-1$.

3. Experiments and Conclusions

To investigate the accuracy of the proposed pitch extraction method, we compared the relative performance of some pitch extraction methods. The methods under evaluation included the cepstrum method (CEP), the AMDF method (AMDF), the autocorrelation function method (ACF), the AMDF-weighted ACF method (AMDF-ACF) [4] and the proposed FWS method (FWS-ACF). Speech data used to evaluate the proposed weighting scheme for pitch extraction was collected over the public telephone network and each Mandarin word comprised of 8 ~ 14 syllables (about 6 ~ 10 seconds). We chose 1000 Mandarin words uttered by 25 males and 25 females for conducting experiments. Each speech frame, which contained 256 samples and 128 samples overlapped, was recorded with the sampling rate of 8 kHz and multiplied by a 256-point Hamming window. For generating noisy speech, additive white Gaussian noise (AWGN) was directly added to the collected telephone speech data in time domain with specific SNR values at ∞ dB, 20 dB, 10 dB and 0 dB. The pitch extraction accuracy was evaluated in terms of the pitch frequency difference between the actual pitch frequency $F_{APF}(n)$ and the estimated pitch frequency $F_{EPF}(n)$, that is [4]-[5]

$$Error(n) = \left| F_{APF}(n) - F_{EPF}(n) \right|. \tag{5}$$

If $Error(n)$ is greater than 10 Hz, the error was treated as a gross pitch error (GPE), and its proportion was denoted as gross pitch error rate (GPER). In Table 1, we demonstrate the influence of fuzzy factor

F on the gross pitch error rates for AWGN. It shows that the GPER initially decreases with the fuzzy factor, attains a minimum value, and then increases with an increase in the fuzzy factor. Obviously, the optimal value of fuzzy factor is related to SNR value, i.e., the lower the SNR value, the smaller the optimal value of fuzzy factor. In addition, the comparisons of GPERs for various methods with AWGN are also listed in Table 2. The experimental results show that by properly adjusting the fuzzy factor, the ACF-based fuzzy weighting scheme can provide higher capability in extracting the pitch information than the AMDF-weighted ACF method and the other conventional approaches we evaluated in noisy environments.

Acknowledgement

This research has been partially sponsored by the National Science Council, Taiwan, R.O.C under contract NSC-90-2213-E-131-001.

REFERENCE

1 HESS, W. J. Pitch Determination of Speech Signals. 1983, Berlin, Germany : Springer-Verlag

2 FURUI, S. Digital Speech Processing, Synthesis, and Recognition. 1989, Marcel Dekker, Inc.

3 OH, K. A., and UN, C. K. : 'A performance comparison of pitch extraction algorithms for noisy speech', *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing.,* 1984, pp. 18B4.1-18B4.4

4 TETSUYA, S., and HAJIME, K. : 'Weighted autocorrelation for pitch extraction of noisy speech', *IEEE Trans., 2001,* SAP-9*, No. 7,* pp. 727-730

5 RABINER, L. R., CHENG, M. J., ROSENBERG, A. E., and MCGONEGAL, C. A., 'A comparative performance study of several pitch detection algorithms', *IEEE Trans.,* 1976, Acoustic., Speech, Signal Processing, Vol. ASSP-24, No. 5, pp. 399-417

Table 1 : Influences of fuzzy factor F on gross pitch error rates (GPERs) (%) for AWGN.

F values	gross pitch error rates (GPERs) (%)			
	NO AWGN	20 dB AWGN	10 dB AWGN	0 dB AWGN
1.01	18.2	35.5	49.8	53.2
1.1	13.4	23.4	32.2	39.5
1.2	10.4	18.2	21.6	28.6
1.3	8.3	13.8	14.9	20.1
1.4	6.4	9.6	10.5	21.8
1.5	4.8	6.5	10.6	24.6
1.6	3.7	5.6	10.9	28.9
1.7	3.1	5.9	11.6	33.1
1.8	3.8	6.4	12.4	35.3
2.0	5.4	7.2	13.3	38.4
10.0	7.4	10.5	15.1	43.2
100.0	8.9	11.8	16.8	47.6

Table 2 : Comparisons of gross pitch error rates (GPERs) (%) for various methods with AWGN.

Methods	gross pitch error rates (GPERs) (%)			
	NO AWGN	20 dB AWGN	10 dB AWGN	0 dB AWGN
CEP	2.9	11.2	26.1	48.3
AMDF	3.3	9.9	20.9	37.6
ACF	3.0	7.3	16.8	29.8
AMDF-ACF	3.1	6.1	12.4	23.6
FWS-ACF	3.1 ($F = 1.7$)	5.6 ($F = 1.6$)	10.5 ($F = 1.4$)	20.1 ($F = 1.3$)