Pitch Detection of Speech Signal Using Wavelet Transform

Amin Shadravan Lalezari, Jalil Shirazi
Communication Engineering Department, Khavaran Higher Education Institute, Mashhad, Iran
aminshadravan@yahoo.com

Abstract: Pitch frequency is the fundamental frequency of a speech signal. It is one of the most important parameters for speech signal processing. The simulated results on Keele pitch reference database show that the performance of the proposed wavelet transform based pitch detection algorithm is obviously better than the original AMDF and its improvements based algorithms.

Keywords—Pitch Detection, Speech Signal, Wavelet Transform, AMDF

I. Introduction

Pitch is one of the most important parameters for speechsignal processing including speech synthesis, automatic speech recognition, speech enhancement etc. Thus it is very important to extract the pitch from the speech accurately.Recently there are many pitch detection methods [1] [2] [3] [4] having been proposed.

During each period of voiced speech the glottis is excited and a GCI (Glottal Closure Instant) occurs. This phenomenon corresponds to a zero crossing in the waveform. If a speech signal is filtered by a derivative function, a maximum will occur at each zero crossing in the waveform. Pitch period detection algorithms are generally divided in two categories; event detection and non-event detection. Event detection algorithms based on autocorrelation function use the relatively prominent peaks in autocorrelation. They have a short coming in estimating pitch period just for a certain vowel, therefor; their efficiency is reduced where speech signal is non-stationary. In non-event detection methods pitch period for a segment of speech signal is calculated by some methods such as cepstrom or average magnitude difference function (AMDF).However, a falling trend presents as a global feature [5] inAMDF so that some detection errors are often happened. It is that the estimated pitch is half or multiple of the actual. To avoid these errors, some improvements of the conventional AMDF were proposed in these literatures [5][6]. These improvements are mainly made that modifying the definition of AMDF (such as CAMDF [5]) or adjusting the length of the frame which is used to compute AMDF (such as EAMDF [6]) to improve the performance of AMDF. Also a new modified AMDF based onEmpirical Mode Decomposition (EMD) [7] to estimate pitch is not very satisfied and will bring other unexpected errors. These methods determine pitch period by a direct approach therefore they are less computation intensive when they operate on windowed speech. Hence they are not suitable for a wide range of speech sources.

During last few years wavelet transform has been used as a tool to analyze many kinds of problems. Kadambe showed when a GCI happens in speech signal, there would be coincident local maximums in its wavelet coefficients for consecutive scales [8]. Therefore pitch period estimation by means of wavelet transform is done by determining the GCI's and measuring the elapsed time between such two adjacent points.

In this paper, we propose a new method based on wavelet transform to estimate pitch period and a high accuracy is ensured at the same time.

The rest of paper is organized as following: Section 2 reviews AMDF, CAMDF, EAMDF and EMDAMDF. After that a pitch detection algorithm based on wavelet transform is proposed. Section 3 gives results of the compared experiments and discussions. Finally, the paper is concluded in Section 4.

II. Material and Methodology

A. Review of AMDF and Its Improvements

The conventional AMDF was proposed by Ross et al. in 1974[2] and it is defined as follows:

$$D(\tau) = \sum_{n=0}^{N-\tau-1} \left| x(n) - x(n+\tau) \right| \qquad (1)$$

Where $x(n)$ denotes a voiced speech frame multiplied by a rectangular window of length N, and τ denotes the lag number.

As shown in Fig. 1(b), instead of true pitch, we estimate a double pitch from AMDF. In this figure, speech is a female voiced frame (Fig. 1(a)) [9].

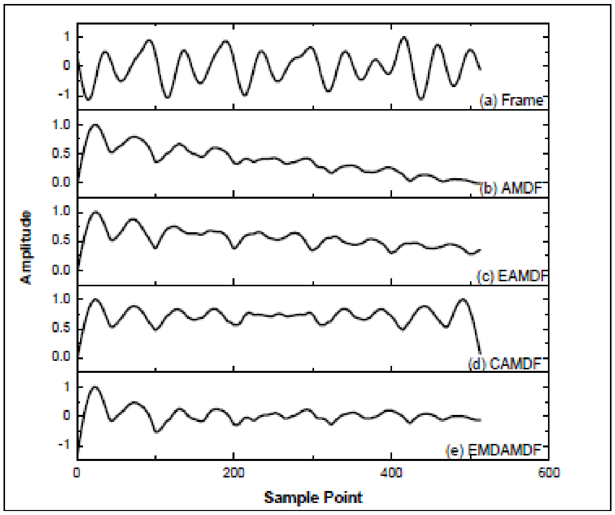

Fig. 1: Comparison between (b) AMDF, (c) EAMDF, (d) CAMDF, and (e) EMDAMDF of (a) a female voiced speech frame [9].

In order to overcome the falling trend of AMDF, CircularAMDF (CAMDF) was proposed in [5] and the description of
CAMDF is given by:

$$D_C(\tau) = \sum_{n=0}^{N-1} \left| x(mod(n+\tau,N) - x(n)) \right| \qquad (2)$$

Where mod (n + τ,N) represents the modulo operation,meaning that (n + τ) modulo$N$.
FromFig. 1(d), we can see that CAMDF eliminates the falling trend, but doublepitch error is still occurred.
In [6], extended AMDF was proposed and high accuracywas reported. EAMDF is defined as following:

$$D_e(\tau) = \frac{1}{N-\tau} \sum_{n=-\frac{N}{2}}^{N+\frac{N}{2}-\tau} \left| x(n) - x(n+\tau) \right| \qquad (3)$$

EAMDF can conquer the falling trend of AMDF.
Fig. 1(c) shows EAMDF of the same speech frame. We can see that double error cannot be conquered.
Empirical mode decomposition AMDF was proposed in [9]. EMDAMDF is defined as following:

$$S_{EMDAMDF}\,(t) = \sum_{n=1}^{N} C_n(t) \qquad (4)$$

In contrast of the original AMDF, EMDAMDF eliminates thefalling trend efficiently and adaptively by using EMD. It can be seen in Fig. 1 that EMDAMDF (Fig. 1(e)) detect the pitch period.

B. Wavelet Transform

The wavelet transform (WT) could be classified as either continuouswavelet transform or discrete wavelet transform(DWT). Acontinuous wavelet transform of a signal$x(t) \in L^2R$ resultsin:

$$WT_x(\omega,\tau) = \frac{1}{\sqrt{\omega}} \int_{-\infty}^{+\infty} x(t)\varphi^*\left(\frac{t-\tau}{\omega}\right) dt \qquad \omega > 0 \qquad (5)$$

Where the function$\varphi(t)$ is usually referred to as mother wavelet, $\omega$is the scaling factor, τ is the shift and $*$ stands for complex conjugation. The DWT can be performed via the multi resolution analysis wavelet decomposition/reconstruction algorithm developed by Mallat. At the m^th level, the multi resolution space, V_m, is spanned by the basic functions$\left\{2^{\frac{m}{2}}\varphi(2^m t - n);\ \ n \in Z\right\}$and the space, W_m, orthogonal to V_m in V_{m-1} is spanned by$\left\{2^{\frac{m}{2}}\emptyset(2^m t - n);\ \ n \in Z\right.$, where $\varphi(t)$is called the scaling function and $\emptyset(t)$is called the wavelet function. Mallat's algorithm allows wavelet coefficients (also called the detailed version of the signal),$d_{m,n} = \langle x(t), \emptyset_{m,n} \rangle$ and scaling coefficients (also called the approximation version)$x_{m,n} = \langle x(t), \varphi_{m,n} \rangle$at the m^th scale to be calculated recursively from the representation of the signal, $x(t)$ at the preceding, finer scale,$x_{m-1,n}$ through the following filtering operation:

$$x_{m,n} = \sum_k a_0(k-2n)x_{m-1,k} \qquad (6)$$

$$d_{m,n} = \sum_k a_1(k-2n)x_{m-1,k} \qquad (7)$$

Where$a_0(n) = \langle \varphi_{1,0}, \varphi_{0,n} \rangle$, $a_1(n) = \langle \emptyset_{1,0}, \emptyset_{0,n} \rangle$.

C. Proposed Pitch Detection Algorithm Based on DWT

First, the segmentation is done by windowing the original signal with a length equal to an approximate duration of a phoneme (i.e. 26.5ms), and jumping of 10ms from each window to the next is employed. Then the wavelet transform of each segment is calculated in 2, 3, 4 and 5 consecutive scales.

After carrying out the above procedure, the local maximums of wavelet coefficients that have a value greater than 70% of the global maximum of the segment are chosen. Among these local maximums of wavelet coefficients, if the distance between the locations of each consecutive local maximums of the segment is less than the lowest pitch period in speech signal (i.e.3ms), the location of the local maximum with higher amplitude is chosen and another one eliminates.
If the locations of these extracted local maximums are the same for at least two consecutive scales, then the segment is considered to be of voiced type, and thepitch period is obtained by measuring the distance between these local maximums. In situation where the locations of these local maximums do not coincide, the segment is considered to be unvoiced, and the pitch period is then taken as zero.
In the present work,Haar wavelet is employed to estimate the pitch period.

III. Results and Tables

We use the *Keele*pitch extraction reference database [10] which is obtained from ftp://ftp.cs.keele.ac.uk/pub/pitch/ to test the performance of the proposed algorithm. Both female (F1-F2-F3) and male (M2-M3-M4) speakers' speech are used here. The speech data is sampled at 20 kHz with 16-bit resolution. The reference pitch values are provided at 100Hz frame rate with 26.5ms rectangular window. Some reference pitch which are recorded as '-1' from the database are manually cut down.
Fig. 2 shows the eligible local maximums of wavelet coefficients of a female voiced speech frame, after carrying out the procedure of proposed method.

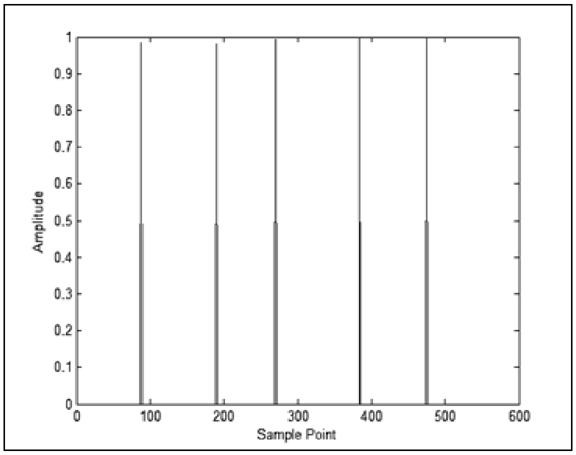

Fig. 2: The eligible local maximums of wavelet coefficients.

We evaluate AMDF, CAMDF, EAMDF, EMDAMDF and the proposed wavelet transform based pitch detection algorithms on *Keele*pitch database. According to the definition of Rabiner [11],if the detected pitch period for a frame defers 1ms from the reference value, the error is defined as a gross pitch error (GPE). The errors are reported in terms of percentage GPE denoted as %GPE.

Table1. Comparisonof Different Algorithms In Terms Of %GPE on Female Speech

	F_1	F_2	F_3
AMDF	22.66	11.93	13.11
CAMDF	9.34	5.73	7.75
EAMDF	7.51	4.58	5.03
EMDAMDF	6.07	3.84	4.63
Proposed Method	**2.64**	**2.90**	**1.67**

Table2. Comparison of Different Algorithms In Terms Of %GPE on Male Speech

	M_2	M_3	M_4
AMDF	9.92	21.31	19.51
CAMDF	7.32	22.04	17.87
EAMDF	3.08	11.33	9.42
EMDAMDF	2.81	9.10	8.35
Proposed Method	**5.52**	**5.52**	**7.39**

As shown in Table1 and Table2, the %GPE of different algorithms for female and male speech are obtained respectively. From these two tables, we can see that the proposed wavelet transform based pitch detection algorithm performs better than all the other functions based algorithms for either female or male, except M_2.

It is also observed that compared with the original AMDF and its improvements, the superiority of the proposed Method can easily be seen on female speech.

IV.Conclusion

In this paper, we give a pitch detection algorithm based on wavelet transform. Finally, a simulated pitch detection experiment based on the *Keele*database is conducted. The results show that the performance of the proposed method based on wavelet transform outperforms the AMDF based improvements such as CAMDF, EAMDF and EMDAMDF in comparison.

Acknowledgement

First thank god, we would like to thank everyone helped us to complete this work.

References

i. M. J. RossH. Shafferand R. Freudberg, et al. "Average magnitude difference function pitch extractor" IEEE Trans. Acoustics, Speech Signal Processing, vol. 22, pp. 353-362, 1974.

ii. M.S. Obaidat, C. Lee, B. Sadoun, D. Nelson. "Estimation of pitch period of speech signal using a newdyadic wavelet algorithm" Information Sciences 119, 21-39, 1999.

iii. ErgunErc,elebi, "Second generation wavelet transform-based pitch period estimation and voiced/unvoiced decision for speech signals" Applied Acoustics 64, 25–41, 2003.

iv. RunshenCai ; Yaoting Zhu ; Shaoqiang Shi. "A Modified Pitch Detection Method Based on Wavelet Transform" Multimedia andInformation Technology (MMIT), 2010 Second International Conference, vol.2, pp. 246 – 249, 2010.

v. W. Zhang, G. Xu, Y. Wang, "Pitch estimation based on circular AMDF" Proceedings of IEEE ICASSP, pp. 341-344, 2002.

vi. Ghulam Muhammad, "Noise Robust Pitch Detection Based on Extended AMDF" Proceedings ofIEEE ISSPIT, pp. 133-138, 2008.

vii. N. E. Huang, S. Zheng, and S. R. Long et al. "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis" Proceedings of Royal Society, pp. 903-995, 1998.

viii. S. Kadambe, G. F. Boudreaux-Bartels, "Application of the Wavelet Transform for pitch detection of speech signals" IEEE Trans. Information Theory, vol. 382, pp. 917-924, 1992.

ix. Yuan Zong; YuminZeng; Mengchao Li; RuiZheng, "Pitch detection using EMD-based AMDF" Intelligent Control and Information Processing (ICICIP), pp. 594-597, 2013.

x. F. Plante. "A pitch extraction reference database" Proceedings of Eurospeech, pp.837-840, 1995.

xi. L. R. Rabiner, M. J. Cheng and C. A. McGonegal, "A comparative performance study of several pitch detection algorithms" IEEE Trans. Acoustics, Speech and Signal Processing, vol. 24, pp. 399-417, 1976.