VT-AMDF, a Pitch Detection Algorithm

Nutthacha Prukkanon[1], Kosin Chamnongthai[1], Yoshikazu Miyanaga[2], and Kohji Higuchi[3]
[1]King Mongkut's University of Technology Thonburi, Bangkok 10140 Thailand
E-mail: nutthacha_pru@utcc.ac.th, kosin.cha@kmutt.ac.th Tel: +662-470-9064
[2]Hokkaido University, Sapporo 060-8628 Hokkaido Japan
E-mail: miya@ist.hokudai.ac.jp Tel/Fax: +81-11-706-6492
[3]Department of Electronic Engineering, University of Electro-Communications, Tokyo, Japan
E-mail: higuchi@ee.uec.ac.jp

Abstract—**This paper proposes a VT-AMDF pitch detection algorithm which based on average magnitude difference function (AMDF). Time processing is important when implement in real time system. In order to decrease computational time and complexity of pitch detection, the VT-AMDF uses only 55% of total time intervals. The method can reduce the computational time 1.3, 1.4, 4.8 and 5.4 times, compared with original AMDF, autocorrelation function (ACF), normalized square difference function (NSDF) and YIN, respectively. The experiments evaluate on seven words in five different tones of Thai isolated word. The results show that the gross error of VT-AMDF is 3.88%, which is more than original AMDF 0.48 and YIN 0.22, but less than ACF 0.54 and NSDF 1.47.**

I. INTRODUCTION

Pitch or fundamental frequency is an essential parameter in tonal languages. The relative pitch motion of an utterance contributes to the lexical information in a word. Speech recognition and speech synthesis attended to pitch to avoid ambiguous meaning. Many pitch detection algorithms have been proposed [1]-[10]. In time domain, pitch detection algorithms are usually computationally simple. The algorithms in frequency domain are usually more complex. The average magnitude different function (AMDF) pitch detection algorithm is performed in time domain and also has been widely applied, especially in real time systems, due to its good accuracy for detection and low complexity [6].

Although many pitch detection algorithms have been proposed, few of them suit for built-in hardware in real time processing. This paper proposes a varied time interval of average magnitude different function algorithm, denoted as VT-AMDF which based on the original AMDF algorithm by varying time interval (τ). The algorithm can reduce the computational time processing by shifting time interval (τ) in different step. It is not necessary to compute the entire τ, since fundamental frequency which derives from portion of sampling rate and τ is not a linear function.

In some pitch detection algorithms, the global minimum point is used for the pitch period. In normal cases, the correct pitch period is the first local minimum point which is also the global minimum point. In some cases, the global minimum is not the first local minimum. Therefore, the pitch detection methods above give a mistake pitch period. In this paper, we present a good detection method that shops around four candidated pitch points. Both doubled/halved pitches will cause pitch detection errors. Thus, smoothing is necessary after pitch detection method. An approach to smooth pitch [11] is employed to enforce continuity of pitch contour.

The VT-AMDF algorithm is evaluated and compared with original AMDF, ACF, YIN and NSDF algorithms. The performance of five pitch detection algorithms were tested on seven words, each word varied with five different tones in Thai isolated word. The experiments show that the proposed algorithm is suitable on both time processing and gross error rate.

The organization of this paper is as follows. Section II describes the five pitch period functions. An approach pitch detection method is presented in section III and pitch smoothing in section IV. Section V details the corpus that uses in experiments and shows results. Finally, the conclusions are provided in Section VI.

II. PITCH PERIOD FUNCTIONS

Five pitch period functions are provided to compare their performance. They are the original average magnitude different function (AMDF) [1], autocorrelation function (ACF), YIN [7], normalized square difference function (NSDF) [4], and the proposed algorithm VT-AMDF.

Speech signal multiplies with Hanning window denoted as s_i. Value of fundamental frequency generally fall in range of 50–300 Hz. So, fundamental frequency between 48-324 Hz is determined, corresponding to search time interval (τ) from 34 to 229 for sampling rate 11 kHz. n is frame number, W is frame length 256 samples and frame shift 128 samples.

A. *Original AMDF [1]*

The original average magnitude different function is defined as

$$\mathrm{AMDF}_n(\tau) = \frac{1}{W}\sum_{i=1}^{W}\left|s_i - s_{i+\tau}\right| \tag{1}$$

AMDF values of a frame are shown in Fig.1 (a).

B. *VT-AMDF (proposed)*

A varied time interval of average magnitude different function (VT-AMDF) based on the original AMDF algorithm by varying time interval (τ). For the original AMDF, each τ increases by one. For the fundamental frequency range 48 to 324 Hz, τ will be varied from 34 to 229 for sampling rate 11 kHz. Therefore, the original AMDF function gives 196 values of each speech frame. When time interval range is high, neighbor fundamental frequencies are slightly difference.

978-1-4244-5016-9/09/$25.00 ©2009 IEEE

Therefore, time interval can step up at high time interval. The idea can reduce the computational time and also decrease the local maxima or minima peak points that cause the pitch detection error. Time intervals are assigned as follows:

$$\mathrm{VT-AMDF}_n(\tau) = \frac{1}{W}\sum_{i=1}^{W}\left|s_i - s_{i+\tau}\right| \qquad (2)$$

where $\tau = \tau_1 + \tau_2 + \tau_3 + \tau_4$

$$\begin{aligned}
\tau_{min} \le \tau_1 &< 0.45\tau_{max} &&; \tau_1 = \tau_1 + 1\\
0.45\tau_{max} \le \tau_2 &< 0.68\tau_{max} &&; \tau_2 = \tau_2 + 2\\
0.68\tau_{max} \le \tau_3 &< 0.93\tau_{max} &&; \tau_3 = \tau_3 + 4\\
0.93\tau_{max} \le \tau_4 &\le \tau_{max} &&; \tau_4 = \tau_4 + 8
\end{aligned}$$

Fig.1 (b) shows the 108 values of VT-AMDF.

C. YIN [7]

YIN is based on the difference function, similar to autocorrelation. The difference function is presented in equation 3.

$$d_t(\tau) = \sum_{i=1}^{W}(s_i - s_{i+\tau})^2 \qquad (3)$$

d(τ) function is zero at zero lag and often nonzero at the period because of imperfect periodicity. In order to reduce the occurrence of subharmonic errors, YIN employed a cumulative mean function which de-emphasized higher-period dips in the difference function. The solution is replaced by $d'_n(\tau)$.

$$d'_n(\tau) = \begin{cases} 1, & \text{if } \tau = 0 \\ d_t(\tau) \Big/ \left[(1/\tau)\sum_{j=1}^{\tau} d_t(j)\right] & \text{otherwise} \end{cases} \qquad (4)$$

$d'_n(\tau)$ values are plotted in Fig.1 (c).

D. ACF

The short time autocorrelation function is obtained by equation 4 and Fig.1(d).

$$\mathrm{ACF}_n(\tau) = \frac{1}{W}\sum_{i=1}^{W} s_i s_{i-\tau} \qquad (5)$$

E. NSDF [10]

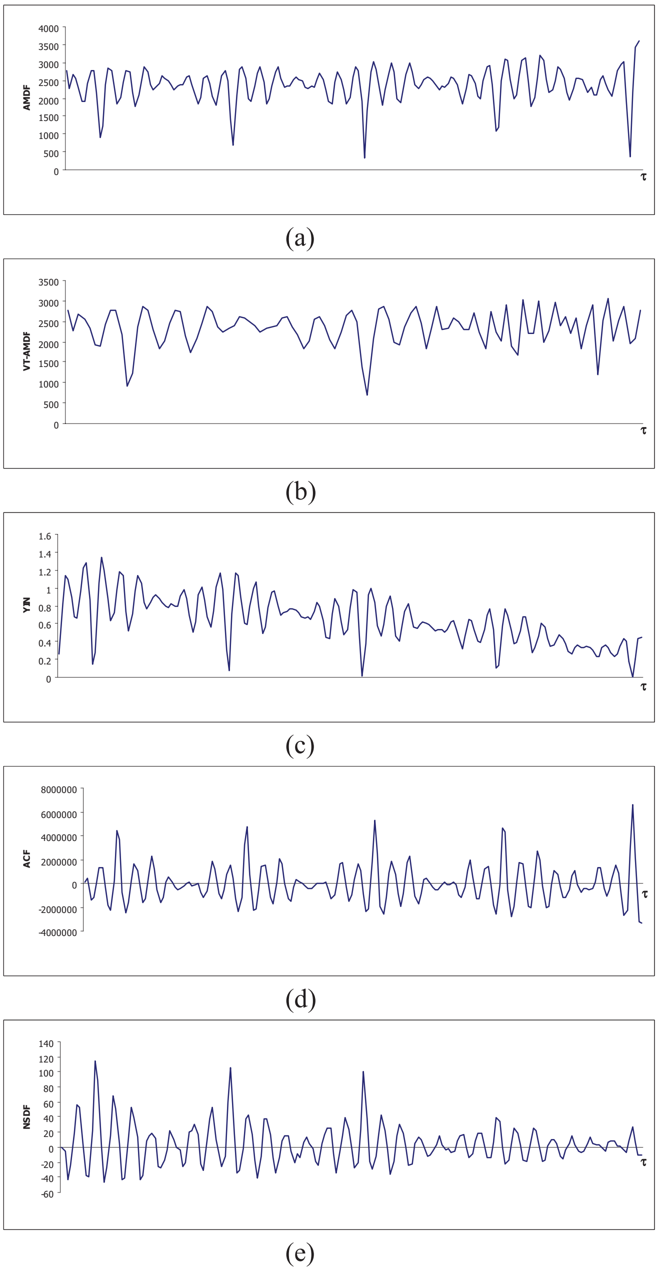
These normalized values simplify the problem of choosing the pitch period as the range is well defined. $\mathrm{NSDF}_n(\tau)$ values more than zero means perfect correlation, equal zero means no correlation and less than zero means perfect negative correlation, as shown in Fig1. (e).

$$\mathrm{NSDF}_n(\tau) = \frac{2r'_n(\tau)}{m'_n(\tau)}$$

$$r'_n(\tau) = \sum_{i=n}^{n+W-1-\tau} s_i s_{i+\tau}$$

$$m'_n(\tau) = \sum_{i=n}^{n+W-1-\tau}(s_i^2 + s_{i+\tau}^2) \qquad (6)$$

Pitch of AMDF, VT-AMDF and YIN are τ that give the first local minimum. Meanwhile, pitch of ACF and NSDF function are τ that give the first local maximum.



(a)

(b)

(c)

(d)

(e)

Fig. 1 Five pitch period functions. (a) AMDF (b) VT-AMDF (c) YIN (d) ACF (e) NSDF

III. Pitch Detection Method

Some pitch detection methods detected pitch period by using threshold. So, pitch period is first τ that gives pitch function more or less than threshold. The threshold is an empirical value from experimental optimization. For the above method, the pitch period might be mistaken by too high or too low threshold value. The other pitch detection methods detected pitch period by using the global minimum or maximum value of pitch period function. In this case, the pitch period might be mistaken as the global minimum or maximum is not the first local minimum or maximum.

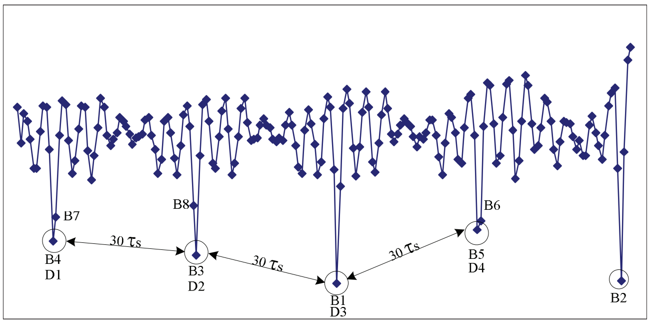


Fig. 2 Pitch detection method

Fig.2 illustrates our pitch detection method, which based on sorting and searching the optimized one of four candidate pitch points. The detection method steps are as follows:

A. Sorting values of pitch period function by ascending.

B. Searching the first eight minimum values of pitch period function, denoted as B1-B8.

C. Sorting the eight minimum values by ascending the τ position.

D. Searching the first four candidate pitch values (D1-D4). The distance of considerate minimum value and the preceding must be more than 30 time intervals (τ).

E. Selecting the one candidate pitch value that is nearest to the preceding frame.

F. If a pitch value differs more than 30 frames of an average pitch value of the preceding and the following, a pitch value will equal to the preceding or following pitch value, depending on direction of checking.

AMDF, VT-AMDF and YIN pitch period function have pitch period at minimum point, mean while, ACF and NSDF have pitch period at maximum point. The above detection method, step A and step B will be changed for ACF and NSDF pitch period function. For step A, sorting by ascending is changed to descending. For step B, searching the minimum values is changed to maximum values.

IV. Pitch Smoothing Algorithm

To eliminate portion of non speech sound, initial consonant and final consonant, Root mean square energy (RMS) is employed for cutting the leading and trailing intervals. The starting point is the first frame that RMS energy is more than threshold, while the end point is searched by inverse direction.

$$\begin{aligned}&\text{if } \left|F_n - F_{n-1}\right| > C_1 \text{ and } \left|F_{n+1} - F_{n-1}\right| > C_2\\&\text{then}\quad F'_n = 2\times F_{n-1} - F_{n-2}\\&\text{if } \left|F_{n+1} - F_{n-1}\right| \le C_2\\&\text{then}\quad F'_n = (F_{n-1} + F_{n+1})/2\end{aligned} \tag{7}$$

where $C_1$and C_2 are thresholds, C_1,C_2=0.1

Some double and halve pitches caused pitch detection errors. Thus, smoothing is necessary after pitch detection method. An approach to smooth pitch [11] is employed to enforce continuity of pitch contour, as equation 6. The error pitches are discarded and then estimated new value from neighbor pitches. Pitch is first normalized and then smoothed in both forward and backward direction.

V. Experiments and Results

A. Data Preparation

The Speech files provided by [12], include 210 words from 6 speakers, 3 males and 3 females in an office environment. The sampling rate for the speech signal is 11 kHz using 16-bit A/D converter. A frame length is 256 samples and 128 sample shifts. Each speaker uttered seven words in five different tones as show in Table I. Speech database was created on empirical studies of effect of an initial consonants, vowels, and final consonants when varying the tones of isolated words.

Table I
List of THAI Isolated Word with Five Tones

pā :	p`a :	pâ :	pa′ :	pă :
nā :	n`a :	nâ :	na′ :	nă :
pī :	p`i :	pî :	pi′ :	pĭ :
pū :	p`u :	pû :	pu′ :	pŭ :
pā : n	p`a : n	pâ : n	pa′ : n	pă : n
pī : n	p`i : n	pî : n	pi′ : n	pĭ : n
pū : n	p`u : n	pû : n	pu′ : n	pŭ : n

B. Pitch Detection Algorithms Evaluation

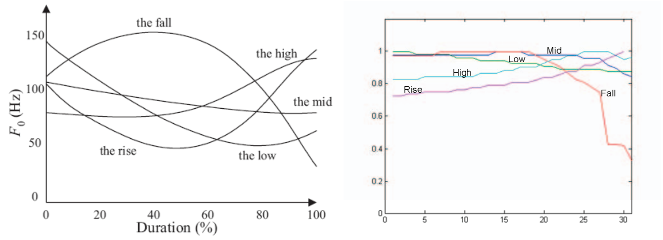
The evaluation is done by calculating the gross error rate of each pitch detection algorithm. True pitch is determined from manual detection by refer to pitch period functions of AMDF, ACF, YIN, and pattern of tone contour. A gross error is counted when estimated pitch is more than 20% higher or lower than true pitch.

Table II
Gross error rates of five PDAs

Pitch Detection Algorithm	Gross error (%)	time/8503 frames (ms)
AMDF	3.40	0.65
VT-AMDF	3.88	0.51
YIN	3.66	2.74
ACF	4.42	0.69
NSDF	5.35	2.43

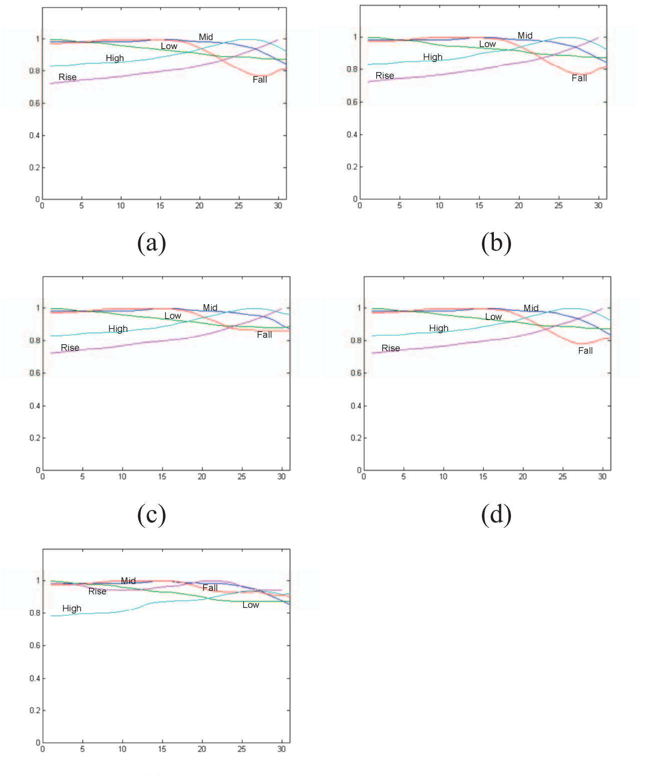
Table II summarizes gross error rates of five pitch detection algorithms. The total voiced frames are 6168, silence is classified from voice in section IV. The experiments show that the error rate of proposed pitch detection algorithm is approximately equal to YIN algorithm, but computational time less than 5.4 times. The method can reduce the computational time 1.3, 1.4 and 4.8 times when compared with AMDF, ACF and NSDF, respectively. (Computer specification: Intel Pentium M processor, 1.73 GHz, 512MB RAM)

Fig. 3 shows the difference of an ideal and true pitch contours of Thai five tones, pā : , p`a : , pâ : , pa′ : , pă : . Since NSDF pitch detection algorithm is designed for musical note sounds, the gross error rate is the most when evaluated on speech sound, as shown in Fig. 4 (e). The others yield almost same performance, but VT-AMDF is the appropriated function in both gross error and computational time. Our evaluation conditions might be different from those they optimized with improved performances. This means that testing on an extension of database is need.



(a) (b)

Fig. 3 pitch contours of five Thai tones (a) ideal pitch [12] (b) true pitch.



(a) (b)

(c) (d)

(e)

Fig. 4 Comparison of five pitch detection algorithms of word pa: with five different tones pā : p`a : pâ : pa′ : pă : (a) pitch contour of original AMDF (b) VT-AMDF (c) YIN (d) ACF (e) NSDF

VI. Conclusions

This paper proposes a VT-AMDF algorithm based on original AMDF. Because the proposed algorithm uses only 55% time interval (τ) of original AMDF function. The algorithm can decrease computational time and complexity when compare with YIN, NSDF. However, the experimental results show that the proposed algorithm is suitable in both time processing and gross error rate, which is simple to implement in hardware.

Acknowledgment

Mrs. Prukkanon would like to express the grateful thanks to University of the Thai Chamber of Commerce (UTCC) for her Ph.D. study and research financial support.

References

[1] Myron J. Ross, Harry L. Shaffer, Andrew Cohen, Richard Freudberg, and Harold J. Manley, "Average Magnitude Difference Function Pitch Extractor", IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-22, pp. 353-362, October, 1974.

[2] Xufang Zhao, Douglas O'Shaughnessy and Nguyen Minh-Quang, "A Processing Method for Pitch Smoothing Based on Autocorrelation and Cepstral F0 Detection Approaches", *2007 International Symposium on Signals, Systems, and Electronics*, pp. 59-62, August 2007.

[3] Li Hui, Bei-qian, and Lu Wie, "A Pitch Detection Algorithm Based on AMDF and ACF", *2006 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. I-377-380, May 2006.

[4] Philip McLeod and Geoff Wyvill, "A Smarter Way to Find Pitch", *Proc. International Computer Music Conference*, pp. 138-141, September 2005.

[5] Li Tan and Montri Karnjanadecha, "Pitch Detection Algorithm: Autocorrelation Method and AMDF", *Proceedings of the 3rd International Symposium on Communications and Information Technology*, vol. 2, pp. 541-546, September 2003.

[6] Yu-Min Zeng, Zhen-Yang Wu, Hai-Bin Liu, and Lin Zhou, "Modified AMDF Pitch Detection Algorithm", *The Second International Conference on Machine Learning and Cybernetice*, pp. 470-473, November 2003.

[7] Alain de Cheveigné and Hideki Kawahara, "YIN, A Fundamental Frequency Estimator for Speech and Music", *Journal of the Acoustical Society of America*, vol. 111(4), pp. 1917-30, April 2002.

[8] Xioa-Dan MEI, Jengshyang Pan, and Sheng-He SUN, "Efficient Algorithms for Speech Pitch Estimation", *2001 International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 421-424, May 2001.

[9] Goangshiuan S. Ying, Leah H. Jamieson, and Carl D. Michell, "A Probabilistic Approach to AMDF Pitch Detection", *The Fourth International Conference on Spoken Language Processing*, vol. 2, pp. 1201-1204, October 1996.

[10] Roudra Chakrabotry, Debapriya Sengupta, and Sagnik Sinha, "Pitch Tracking of Acoustic Signals based on Average Squared Mean Difference Function", *Signal, Image and Video Processing*, Springer London, 2008.

[11] Liu Jun, Xiaoyan Zhu, and Yuping Luo, "An Approach to Smooth Fundamental Frequencies in Tone Recognition", *International Conference on Communication Technology*, pp. S16-10-1-S16-10-5, October 1998.

[12] Nuttakorn Thubthong, "A Study of various linguistic effects on tone recognition in Thai continuous speech", *Ph.D. Thesis, Chulalongkorn University, Bangkok, 2002*.