American Journal of Applied Sciences 9 (1): 93-102, 2012
ISSN 1546-9239

An Improved Time Domain Pitch Detection Algorithm for Pathological Voice

Mohd Redzuan Jamaludin, Sheikh Hussain Shaikh Salleh, Tan Tian Swee, Kartini Ahmad, Ahmad Kamarul Ariff Ibrahim and Kamarulafizam Ismail
Centre for Biomedical Engineering,
University Technology Malaysia Skudai, Malaysia

Abstract: Problem statement: The present study proposes a new pitch detection algorithm which could potentially be used to detect pitch for disordered or pathological voices. One of the parameters required for dysphonia diagnosis is pitch and this prompted the development of a new and reliable pitch detection algorithm capable of accurately detect pitch in disordered voices. **Approach:** The proposed method applies a technique where the frame size of the half wave rectified autocorrelation is adjusted to a smaller frame after two potential pitch candidates are identified within the preliminary frame. **Results:** The method is compared to PRAAT's standard autocorrelation and the result shows a significant improvement in detecting pitch for pathological voices. **Conclusion:** The proposed method is more reliable way to detect pitch, either in low or high pitched voice without adjusting the window size, fixing the pitch candidate search range and predefining threshold like most of the standard autocorrelation do.

Key words: Pitch Detection Algorithm (PDA), dysphonia, autocorrelation, Merged Normalized Forward Backward Correlation (MNFBC), pathological voices, Hilbert-Huang Transform (HHT), time domain, mean error, Auto Correlation Function (ACF)

INTRODUCTION

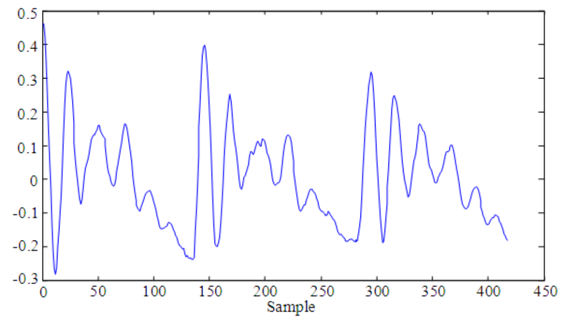
Vocal cords within the laryngeal structure vibrate due to air passing through them during voiced speech (Swee *et al*., 2010). During voiced phonation pitch is produced and the fundamental frequency, F_0 and its reciprocal known as pitch period, T_0 can be calculated (Amado and Filho, 2008; Kotnik *et al*., 2009; Manfredi *et al*., 2000). Vocal hyperfunction, vocal abuse and misuse, or unhealthy social habits such as smoking and alcohol consumption may over time, cause physical changes to the laryngeal structure and lead to voice changes such as loss of power, changes in pitch and reduction in voice range (Hadjitodorov and Mitev, 2002; Timmermans *et al*., 2002; Godino-Llorente *et al*., 2006; De Bodt *et al*., 2007).

Cycle-to-cycle pitch period perturbation (also known as jitter) is usually one of the parameters used to measure voice quality. In order to obtain an accurate pitch period for each cycle of voiced phonation, the Pitch Detection Algorithm (PDA) needs to be able to perform equally well in pathological voices (Manfredi *et al*., 2000; Jang *et al*., 2007; Schoentgen, 2003). The detection of pitch is difficult due to the following reasons:

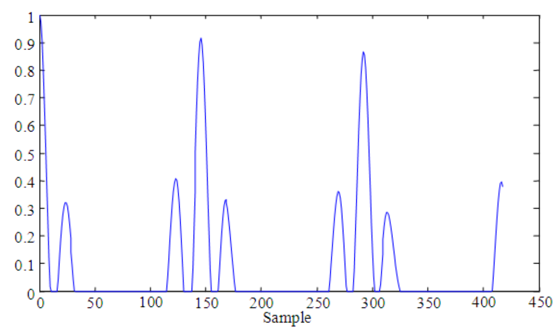
- The nonstationarity and quasiperiodicity of the speech signal as well as the interaction between the glottal excitation and the vocal tract (Ahmadi and Spanias, 1999; Chen and Wang, 2001; Rabiner *et al*., 1976)
- False pitch estimates can also be caused by noise and signal distortion that occur in real environments and errors in voicing decision (Cai and Liu, 1997; Tabrikian *et al*., 2004; Chomphan, 2011)
- For dysphonic voices, there are significant perturbation of amplitude and frequency in the voiced signal, presence of subharmonic and aperiodic components of high intensity and also influence of voiced signal formant structure (Mitev and Hadjitorov, 2003)

Many Pitch Detection Algorithms (PDA) have been developed and yet the results are not adequately reliable in detecting pitch in pathological voices (Mitev and Hadjitorov, 2003).

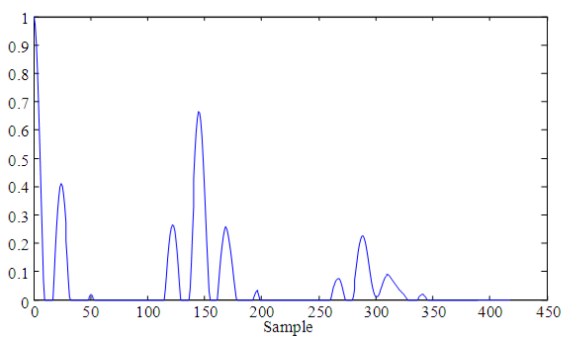
Corresponding Author: Mohd Redzuan Jamaludin, Centre for Biomedical Engineering University Technology Malaysia Skudai, Malaysia



(a)



(b)



(c)

Fig. 1: (a) Acoustic waveform of an /a/ utterance; (b) The corresponding ACF according to Eq. 1; (c) The corresponding ACF according to Eq. 2

This study aimed to propose a newly developed time domain PDA with improved reliability in detecting disordered pitch. The PDA was tested on the KayPENTAX Elemetrics database for the vowel /a/ from 50 normal voices and 100 pathological voices randomly selected. The results were compared with the datasheet provided by KayPENTAX Elemetrics for the accuracy test. The performance of the proposed PDA was also compared with the well-known and publicly available PRAAT toolkit (Kotnik *et al*., 2009).

There are several known types of time domain based PDA. The most prominent one is the Auto Correlation Function (ACF). The following shows the general equation of the ACF (Abdullah-Al-Mamun *et al*., 2009; De Cheveigne and Kawahara, 2002; Quatieri, 2002; Lahat *et al*., 1987; Momani, 2009):

$$R_x = \sum_{n=i}^{i+N-1} s[n]s[n+l] \quad (1)$$

Where:

R_x = The autocorrelation value
s[n] = The input speech signal at sample
i = The first sample inside a frame n
N = The frame size
l = The lag or time displacement that ranges from zero to the number of sample per frame minus one

The lag value that produces maximum peak will be chosen as the pitch period. According to De Cheveigne and Kawahara (2002), another type of autocorrelation equation is as the following:

$$R_x = \sum_{n=i}^{i+N-l-1} s[n]s[n+l] \quad (2)$$

Figure 1a is a frame of speech waveform of the vowel /a/. The equation includes the lag, l to be subtracted from n to produce ACF as shown in Fig. 1c while Eq. 1 produces ACF in Fig. 1b and 2. ACF produced by Eq. 2 degraded as the l value increases by time.

Figure 1b-c show the ACF of the acoustic waveform which were normalized and half-wave rectified from Fig. 1a. It can be seen from Fig. 1b that there are two dominant ACF peaks and these are termed as pitch candidates. The first peak is at lag = 146 and the second peak lies at lag = 292. Usually, to choose the best pitch to be defined as the pitch period of the frame, a rule must be set whereby the range of choosing the best pitch should not be near to zero lag and should not exceed certain value of lag. This rule reflects the limit for human pitch range which is 60-500 Hz (Mitev and Hadjitorov, 2003). Most of the existing commercialized software such as PRAAT and Computerized Speech Laboratory by Kay Elemetrics require the users to specify their own fundamental frequency range of interest in order for the algorithm to work efficiently. Some literature also proposed the use of ACF threshold so that only peaks that exceed this predetermined threshold will be notified as pitch candidates (Mitev and Hadjitorov, 2003). But these rules lack flexibility. If the range is poorly specified, the algorithm will take the wrong lag as pitch period. If the range is not specified at all, the autocorrelation will not be able to accurately detect low pitched voice as reported by Samad *et al*. (2000). The threshold rule will also be inappropriate for Eq. 1-2 since some of the voices' ACF do not even exceed 0.5 or more.

Another method is called the Average Magnitude Difference Function (AMDF) (Manfredi *et al*., 2000). The general equation for AMDF, R_y is as follows Eq. 3 (Quatieri, 2002; Chong and Shih-Chien, 1977):

$$R_y = \sum_{n=i}^{i+N-1} \left| s[n] - s[n+l] \right| \tag{3}$$

Unlike ACF which selects the maximum peak as the pitch candidate, AMDF tends to search the minimum peak as the pitch candidate. Manfredi *et al*. (2000) proposed the modified AMDF where the first valley found to be less than the threshold is set to be the pitch period of the frame. This approach also has its weakness similar to the ACF whereby some harmonics and noise effects can also produce AMDF values that falls below this threshold.

From these basic time domain PDA's, many researchers have modified these algorithms so that it will work more efficiently to obtain pitch. One of the interesting approaches was Merged Normalized Forward Backward Correlation (MNFBC) which basically used the same concept of autocorrelation but instead of using autocorrelation, it uses MNFBC which is to be noise robust (Kotnik *et al*., 2009). Plus the method of finding the exact pitch period was by implementing viterbi search to the MNFBC. The viterbi searches for three largest value of the MNFBC as the pitch candidates per voiced frame. But the viterbi search introduces high dependency on current frame's pitch value with the previous frame's pitch value and it will not be able to work efficiently with dysphonic voices since cycle-to-cycle pitch period can vary extremely from each other. False period estimation can also occur when the MNFBC value is larger at pitch candidates other than the true pitch period.

Huang and Pan (2006) and Donato *et al.* (1999) proposed Hilbert-Huang Transform (HHT) for PDA which was developed to consider the non-linearity characteristics of speech signal. It was proven to produce better accuracy of pitch detected but the computational requirements are also increase (Kotnik *et al*., 2009).

Jang *et al*. (2007) Experimented several PDA's to be implemented on pathological voices and the result showed that ACF was the most credible PDA to detect pathological voice. Mitev and Hadjitorov (2003) presented that with a little modification to ACF, it can be an accurate PDA to be applied to pathological voices. But the method still depends on a threshold which they used was 0.5. Some of the pathological voices have fewer ACF than 0.5 even at the pitch period. These findings indicate that ACF time domain based PDA can still be able to detect pitch in dysphonic voices with high accuracy.

From all of the information given above, this study is proposing ACF with modification and with less computational cost for pitch detection in dysphonic voices without using a predetermined threshold and can also automatically set the pitch searching range unlike most of the commercial software where the users themselves need to set the searching range.

MATERIALS AND METHODS

The proposed algorithm for PDA is based on time domain approach consists of the modified ACF. The procedures are as the following.

Step (1): Initialization: Let t = 1 be the initial point of the speech signal. The frame size used for the algorithm is two times maximum pitch period, MAX_PER. MAX_PER is the lowest pitch that human can produce which is 60 Hz of voiced speech signal so that at least within this frame size, two best ACF peaks can be chosen as pitch candidates.

Step (2): Compute autocorrelation: The autocorrelation equation used is Eq. 2. The equation will produce ACF or R_x with i is equals to t, l ranges from zero to 2*MAX_PER - 1 and N is equals to 2*MAX_PER.

Step (3): Half-wave rectification and normalization: The ACF is then normalized and half-wave rectified so that the values for consideration are normalized and positive. This technique was introduced by Kotnik *et al*. (2009) using the following procedures:

- R_x is calculated using Eq. 2. R_o and R_t are found using the following formulae Eq. 4 and 5:

$$R_o = \sum_{n=i}^{i+N-1} s[n]s[n] \qquad (4)$$

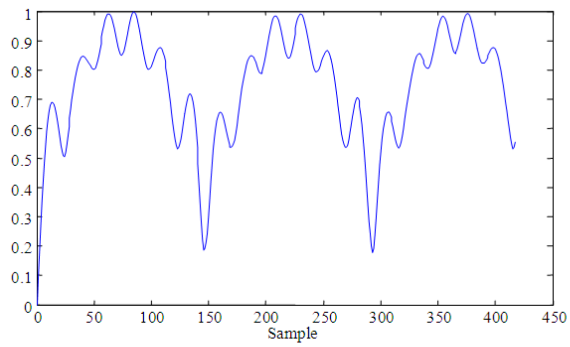
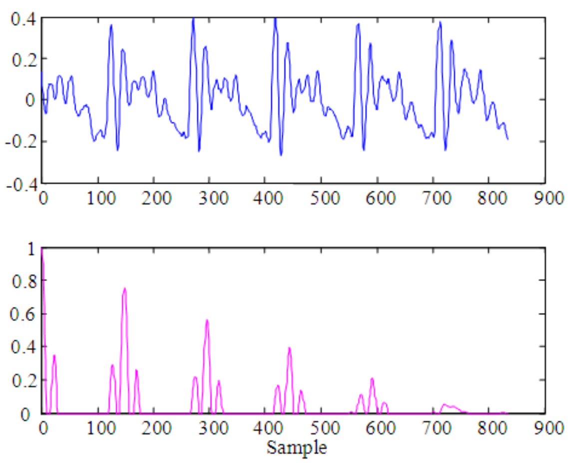
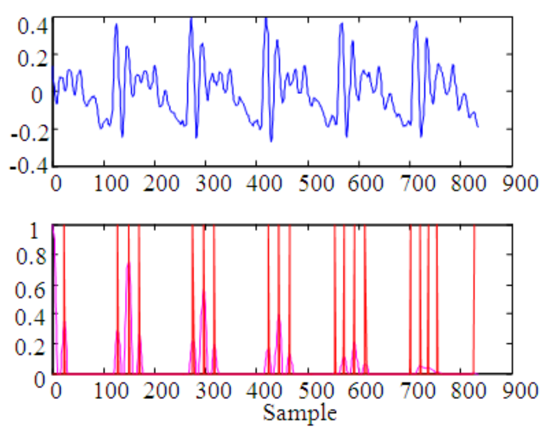


Fig. 2: AMDF of Fig. 1a's waveform



(a)



(b)

Fig. 3: (a) ACF of an /a/ utterance. (b) Marked ACF peaks

$$R_t = \sum_{n=i}^{i+N-1} s[n+l]s[n+l] \qquad (5)$$

where similar to (2), s[n] is the input speech signal at sample n, i is the first sample inside a frame, N is the frame size and l is the lag. Lag l ranges from zero until N-1.

- Then the normalization of R_x is done by using the following formula Eq. 6 and 7:

$$R = \frac{\sum_{n=i}^{i+N-1-l} s[n]s[n+l]}{\sqrt{\left(\sum_{n=i}^{i+N-1} s[n]s[n]\right)\left(\sum_{n=i}^{i+N-1} s[n+l]s[n+l]\right)}} \qquad (6)$$

Or:

$$R = \frac{R_x}{\sqrt{R_o \times R_t}} \qquad (7)$$

- Then R is half-wave rectified by setting all the negative values to zero

Step (4): Mark all possible candidates: All the peaks of the ACF are then marked as possible pitch candidates, $T_i(i)$. Figure 3a shows one frame of ACF of the vowel /a/ and Fig. 3b is the marked peaks, T_i. The algorithm has considered several conditions for the system to work efficiently after every T_i are being recognized:

- If there is no T_i, or i = 0, then the pitch for that frame is set to 0 and the frame moves to the next frame as much as 2*MAX_PER.
- If T_i exists, go to step (5)

Step (5): Find two best candidates: The algorithm will then find the best two candidates by sorting the R_x at every T_i from the largest value to the lowest value along with their T_i as shown in Table 1. From the rearranged candidates, the best two candidates are found by firstly use the following Eq. 8 to find the difference between a pair of T_i:

$$\text{diff} = |T_i(1) - T_i(j)| \qquad (8)$$

where, j = 2, 3, 4,…, j_{total} and the best two candidates are chosen based on the following condition Eq. 9:

$$\text{diff} \geq (2 * \text{MAX\_PER}) / 8 \qquad (9)$$

Table 1: Marked peaks before rearrangement and after rearrangement of ACF shown in Fig. 3

	Marked candidates		Rearranged according to descending order of ACF values	
j	ACF (R)	Period (T_{i1})	ACF (R)	Period (T_{i1})
1	0.3472	22	0.75920	149
2	0.2931	128	0.56330	296
3	0.7592	149	0.39260	443
4	0.2607	170	0.34720	22
5	0.2209	275	0.29310	128
6	0.5633	296	0.26070	170
7	0.1935	317	0.22090	275
8	0.1721	423	0.21460	592
9	0.3926	443	0.19350	317
10	0.1341	465	0.17210	423
11	0.0106	553	0.13410	465
12	0.1159	571	0.11590	571
13	0.2146	592	0.06910	613
14	0.0691	613	0.05580	721
15	0.0130	701	0.04040	737
16	0.0558	721	0.01390	754
17	0.0404	737	0.01300	701
18	0.0139	754	0.01060	553
19	0.0021	826	0.00216	826

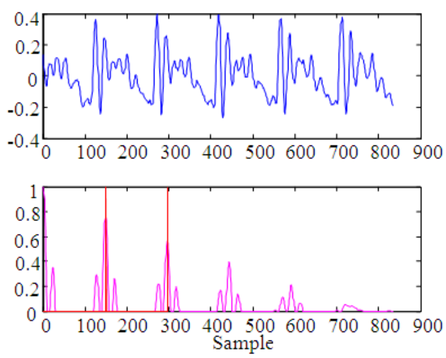


Fig. 4: The marks indicate the best two candidates chosen

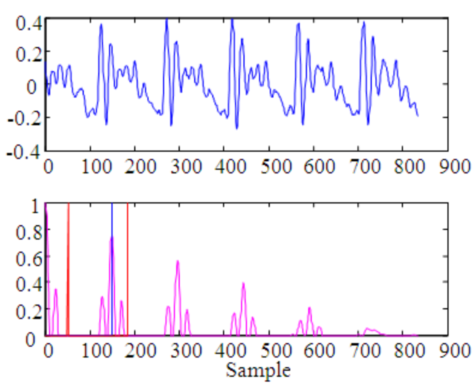


Fig. 5: The red marks indicate the range between new_framei and new_framef. The blue line marks the pitch period

The first T_i pair that achieves this condition will be kept as b_1 and b_2 for the next step. The value (2*MAX_PER)/8 was obtained experimentally as values lower or higher than this will degrade the performance of the proposed PDA which is to accurately detect the pitch. Figure 4 shows the two pitch candidates which have been marked.

Step (6): Create new frame: Once the two candidates are selected, the size of the new frame will be calculated as the following Eq. 10-12:

$$c = |b_1 - b_2| \tag{10}$$

$$new\_framei = MIN\_PER \tag{11}$$

$$new\_framef = c + (c/4) \tag{12}$$

Where:
new_framei = The initial point of the new frame while
new_framef = The final point of the new frame

Instead of searching the pitch within 1 until MAX_PER-1 range or within a predefined range as most of the ACF does in the literature, this study introduces the new searching range which will be from new_framei until new_framef. With the new frame introduced, the largest $R_x(T_i)$ value that lies within that range will be chosen and its corresponding T_i is considered as the pitch period, T_0. Figure 5 shows the new frame or the new region to search the T_0 and the T_0 is marked with blue line.

Step (7): Proceed to the next frame: Since the frame size used might be consisting of two or more pitches, the starting point of the new frame is found according to the following Eq. 13:

$$t_{new} = t_{prev} + T_0 \tag{13}$$

Where:
t_{new} = The new frame's starting point,
t_{prev} = The previous frame's starting point and
T_0 = The pitch period found from the previous frame

This way, every pitch period or every pitch epoch can be located accurately as shown in Fig. 6.

The experiment was conducted to test the accuracy and the effectiveness of the proposed PDA on normal voices and pathological voices.

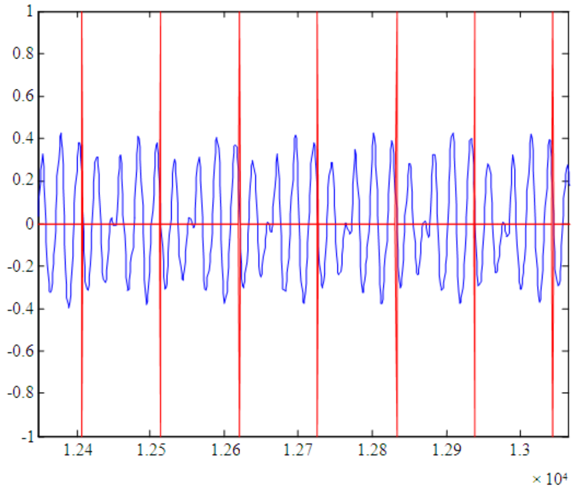


Fig. 6: The marks indicate the start and the end of a pitch period

One of the experiments conducted was applying the PDA on /a/ utterances from the KayPENTAX Elemetrics voice database consists of 50 normal voices and 100 functional and organic voice disorders. The parameters that were used to be compared with the datasheet of results given along with the database were the average fundamental Frequency (F0), highest Fundamental frequency (Fhi), lowest fundamental frequency (Flo) and Standard deviation of the fundamental frequency (STD). The reference values were considered as the true values. The error percentage was calculated by using the following Eq. 14:

$$\text{err}(\%) = \frac{\left|p_{\text{proposedPDA}} - p_{\text{reference}}\right|}{p_{\text{reference}}} \times 100\% \qquad (14)$$

Where:

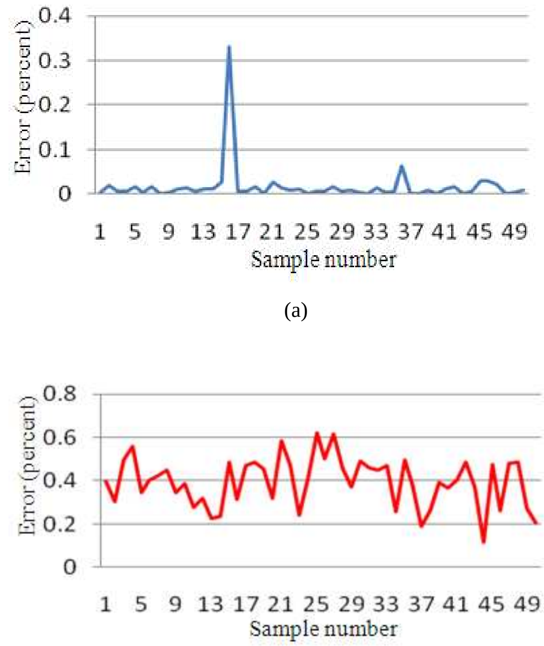
$p_{\text{proposedPDA}}$ = The value of each parameter obtain by using the proposed PDA

$p_{\text{reference}}$ = The value of each parameter given by the reference

The results were also compared with the well-known and publicly available PRAAT toolkit where the PRAAT autocorrelation (PRAAT_ac) was chosen because the proposed algorithm is a modified autocorrelation (Kotnik *et al*., 2009).

RESULTS

Table 2 shows the errors of each parameter produced by using the proposed PDA while Table 3 presented the errors of each parameter produced when PRAAT_ac was used.



(a)

(b)

Fig. 7: (a) Error percentage for mean fundamental frequency of PRAAT (autocorrelation) for normal voice. (b) Error percentage for mean fundamental frequency of proposed PDA for normal voice

Figure 7-10 shows the comparison of the error produced by using PRAAT_ac and the proposed PDA. The observation shows that the PRAAT_ac works well for normal voices as to compare with the proposed algorithm. However, the error differences between PRAAT_ac and the proposed PDA are only at a very small scale.

Table 2-3 show the mean of the errors for each voice sample and each parameter by using two different PDA's. According to the results from Table 2-3, PRAAT_ac produces more error for the pathological voice than the proposed PDA.

To summarize the result obtained by using the proposed PDA and PRAAT_ac, every parameter was averaged to get the mean error for each voice sample. For the proposed PDA it has been found that for 49 normal voices, the mean error was less than 20% and one voice was classified to be having more than 20% error, 13 pathological voices had more than 20% average error while another 87 pathological voices had less than 20% mean error. These data are presented in Table 4.

Table 2: Errors in percentage produced by using the proposed pitch detection algorithm

Voice pattern	Normal	Pathological
Mean fo	0.394	1.189
Highest fo	1.938	4.695
Lowest fo	1.260	4.560
Standard deviation	21.713	35.915

Table 3: Errors in percentage produced by using PRAAT (autocorrelation)

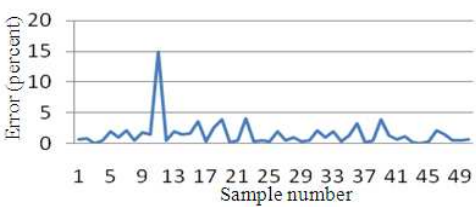
Voice pattern	Normal	Pathological
Mean fo	0.016	1.779
Highest fo	1.508	5.555
Lowest fo	1.375	5.473
Standard deviation	20.423	56.633

Table 4: Classification of voices according to the mean error of each voice sample using the proposed pitch detection algorithm

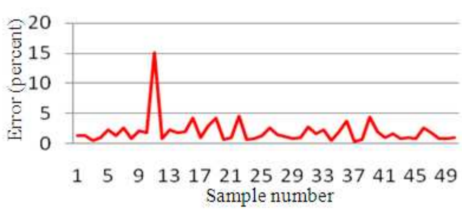
Voice pattern	Error < 20%	Error > 20%
Normal	49	1
Pathological	87	13

Table 5: Classification of voices according to the mean error of each voice sample using the praat_ac

Voice pattern	Error < 20%	Error > 20%
Normal	49	1
Pathological	85	15



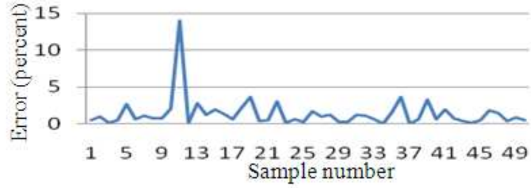
(a)



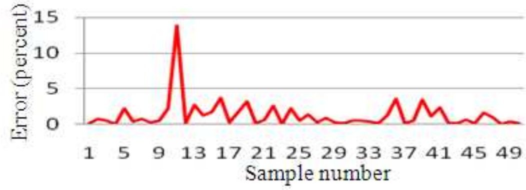
(b)

Fig. 8: (a) Error percentage for highest fundamental frequency of PRAAT (autocorrelation) for normal voice; (b) Error percentage for highest fundamental frequency of proposed PDA for normal voice

While in Table 5, similar to the proposed PDA, there were 49 voices identified to be having less than 20% error and only one voice was put in the more than 20% error category. For pathological voice, there are 15 voices were having mean error of more than 20% and another 85 voices had less than 20% mean error.

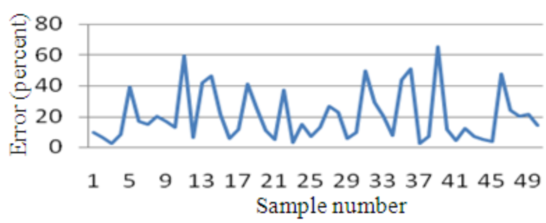


(a)

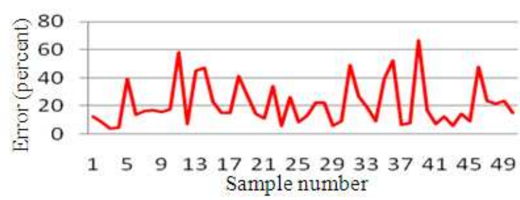


(b)

Fig. 9: (a) Error percentage for lowest fundamental frequency of PRAAT (autocorrelation) for normal voice. (b) Error percentage for lowest fundamental frequency of proposed algorithm for normal voice



(a)

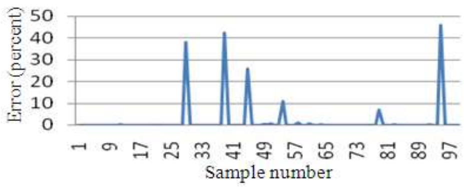


(b)

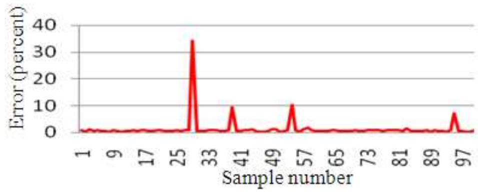
Fig. 10: (a) Error percentage for standard deviation of the fundamental frequency of PRAAT (autocorrelation) for normal voice. (b) Error percentage for standard deviation of the fundamental frequency of proposed algorithm for normal voice

DISCUSSION

Even though it was observed that PRAAT_ac works better for normal voices, Figure 11 until Fig. 14 presented that it works poorly for pathological voices while the error produced by using the proposed algorithm is smaller for pathological voices.

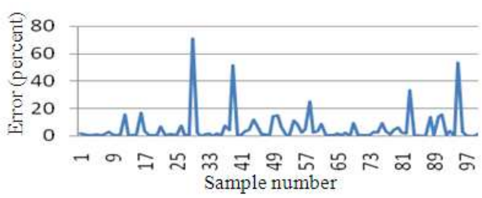


(a)

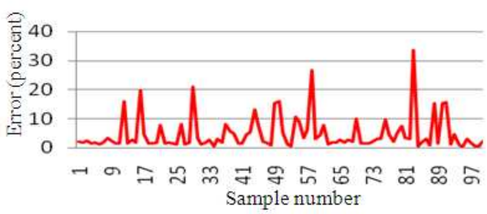


(b)

Fig. 11: (a) Error percentage for mean fundamental frequency of PRAAT (autocorrelation) for pathological voice. (b) Error percentage for mean fundamental frequency of proposed algorithm for pathological voice



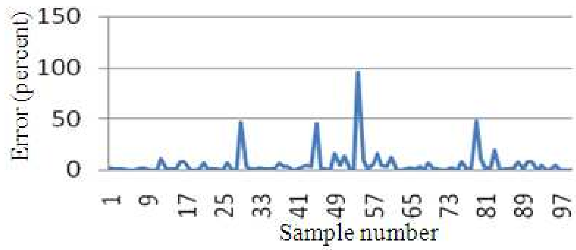
(a)



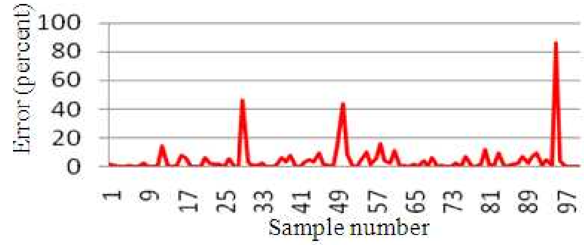
(b)

Fig. 12: (a) Error percentage for highest fundamental frequency of PRAAT (autocorrelation) for pathological voice. (b) Error percentage for highest fundamental frequency of proposed algorithm for pathological voice

Figure 12 shows that three samples exceed 40% of error by using PRAAT_ac while the proposed algorithm had no error that exceeds 40% of error. Figure 13 also indicates that proposed algorithm produces less error by having three samples with more than 40% of error while the PRAAT_ac produces four samples with more than 40% of error. As can be seen in Fig. 14, the standard deviation error for PRAAT_ac exceeds 100% for two sample pathological voices.

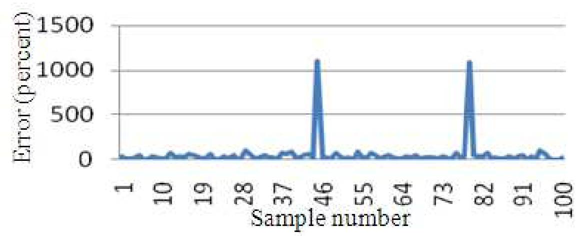


(a)

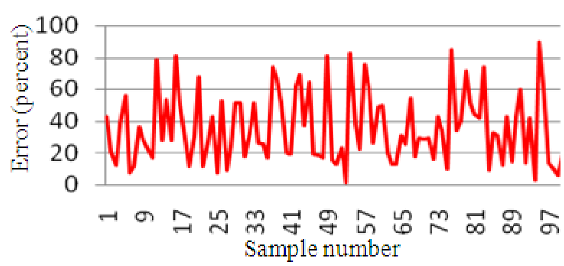


(b)

Fig. 13: (a) Error percentage for lowest fundamental frequency of PRAAT (autocorrelation) for pathological voice. (b) Error percentage for lowest fundamental frequency of proposed algorithm for pathological voice

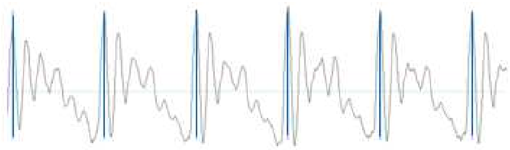


(a)



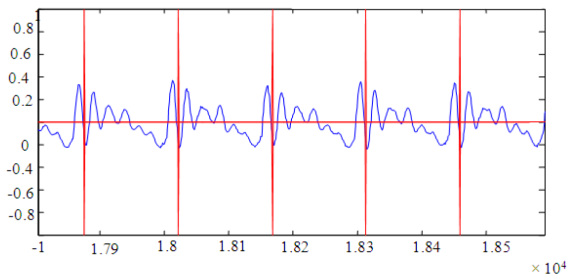
(b)

Fig 14: (a) Error percentage for standard deviation of the fundamental frequency of PRAAT (autocorrelation) for pathological voice. (b) Error percentage for standard deviation of the fundamental frequency of proposed algorithm for pathological voice



(a)

(b)



(c)

Fig. 15: (a) PRAAT analysis window showing marks (blue line) of pitch period detected for the first part of the pathological voice signal sample number 44. (b) PRAAT analysis window showing marks (blue line) of pitch period detected for the second part of the pathological voice signal sample number 44. (c) The red line marks indicate each pitch period by using the proposed algorithm for the pathological voice signal sample number 44

Figure 15 shows the window of PRAAT_ac marking the pitch period of a pathological signal sample number 44.

It can be seen in Fig. 14 that the error of the standard deviation of sample number 44 is over 1000% as well as sample number 78. As can be observed in Fig. 15a, the PRAAT_ac marked the pitch period correctly for the first half of the signal but marked the pitch period wrongly for the second half of the signal shown in Fig. 15b. This is maybe due to the autocorrelation used for the pitch detecting whereby the second pitch period has higher ACF value than the first or the true pitch period. This will happen if the search criterion for the autocorrelation only involves finding the maximum ACF within a predefined range.

But by implementing the proposed algorithm to the same voice sample as can be seen in Fig. 15c, the pitch period can be well determined along the signal thus producing a smaller error than PRAAT_ac.

In both methods, the voice samples with error of more than 20% are due to the strong subharmonics frequencies. The disordered voice with creaky or breathy characteristics will also influence the signal's waveform and since the autocorrelation is dependent upon the signal's amplitude and how correlate the periodic pattern is, the autocorrelation function produced will also be distorted.

CONCLUSION

The proposed method of determining pitch provides significant improvement to the standard autocorrelation which in this case is indicated by the autocorrelation by PRAAT for disordered voice. It allows a more reliable way to detect pitch, either in low or high pitched voice without adjusting the window size, fixing the pitch candidate search range and predefining threshold like most of the standard autocorrelation do.

REFERENCES

Abdullah-Al-Mamun, K., F. Sarker and G. Muhammad, 2009. A high resolution pitch detection algorithm based on AMDF and ACF. J. Sci. Res., 1: 508-515. DOI: 10.3329/jsr.v1i3.2569

Ahmadi, S. and A.S. Spanias, 1999. Cepstrum-based pitch detection using a new statistical V/UV classification algorithm. IEEE Trans. Speech Audio Process., 7: 333-338. DOI: 10.1109/89.759042

Amado, R.G. and J.V. Filho, 2008. Pitch detection algorithms based on zero-cross rate and autocorrelation function for musical notes. Proceeding of the International Conference on Audio, Language and Image Processing, Jul. 7-9, IEEE Xplore Press, Shanghai, pp: 449-454. DOI: 10.1109/ICALIP.2008.4590188

Cai, J. and Z.Q. Liu, 1997. Robust pitch detection of speech signals using steerable filters. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Apr. 21-24, IEEE Xplore Press, Munich, Germany, pp: 1427-1430. DOI: 10.1109/ICASSP.1997.596216

Chen, S.H. and J.F. Wang, 2001. Extraction of pitch information in noisy speech using wavelet transform with aliasing compensation. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 7-11, IEEE Xplore Press, Salt Lake City, UT, USA., pp: 89-92. DOI: 10.1109/ICASSP.2001.940774

Chomphan, S., 2011. Effects of noises on the analysis of fundamental frequency contours for Thai speech. J. Comput. Sci., 7: 568-572. DOI: 10.3844/jcssp.2011.568.572

Chong, U. and Y. Shih-Chien, 1977. A pitch extraction algorithm based on LPC inverse filtering and AMDF. IEEE Trans. Acoust. Speech Signal Process., 25: 565-572. DOI: 10.1109/TASSP.1977.1163005

De Bodt, M.S., K. Ketelslagers, T. Peeters, F. L. Wuyts and F. Mertens *et al*., 2007. Evolution of vocal fold nodules from childhood to adolescence. J. Voice, 2: 151-156. DOI: 10.1016/j.jvoice.2005.11.006

De Cheveigne, A. and H. Kawahara, 2002. YIN, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Am., 111: 1917-1930.

Donato, G.. M.S. Bartlett, J.C. Hager, P. Ekman and T.J. Sejnowski, 1999. Classifying facial actions. IEEE Trans. Patt. Anal. Mach. Intell., 21: 974-989. DOI: 10.1109/34.799905

Godino-Llorente, J.I., N. Soenz-Lechon, V. Osma-Ruiz, S.Aguilera-Navarro and P. Gomez-Vilda, 2006. An integrated tool for the diagnosis of voice disorders. Med. Eng. Phys., 28: 276-289. DOI: 10.1016/j.medengphy.2005.04.014

Hadjitodorov, S. and P. Mitev, 2002. A computer system for acoustic analysis of pathological voices and laryngeal diseases screening. Med. Eng. Phys., 24: 419-429. DOI: 10.1016/S1350-4533(02)00031-0

Huang, H. and J. Pan, 2006. Speech Pitch Determination based on Hilbert-Huang transform. Signal Process., 86: 792-803. DOI: 10.1016/j.sigpro.2005.06.011

Jang, S.J., S.H. Choi, H.M. Kim, H.S. Choi and Y.R. Yoon, 2007. Evaluation of performance of several established pitch detection algorithms in pathological voices. Proceeding of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 22-26, IEEE Xplore Press. Lyon, pp: 620-623. DOI: 10.1109/IEMBS.2007.4352366

Kotnik, B., H. Hoge and Z. Kacic, 2009. Noise robust f0 determination and epoch-marking algorithms. Signal Process., 89: 2555-2569. DOI: 10.1016/j.sigpro.2009.04.017

Lahat, M., R. Niederjohn and D. Krubsack, 1987. A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech. IEEE Trans. Acoust. Speech Signal Process., 35: 741-750. DOI: 10.1109/TASSP.1987.1165224

Manfredi, C., M. D'Aniello, P. Bruscaglioni and A. Ismaelli, 2000. A comparative analysis of fundamental frequency estimation methods with application to pathological voices. Med. Eng. Phys., 22: 135-147. DOI: 10.1016/S1350-4533(00)00018-7

Mitev, P. and S. Hadjitorov, 2003. Fundamental frequency estimation of voice of patients with laryngeal disorders. Inform. Sci.. 156: 3-19. DOI: 10.1016/S0020-0255(03)00161-0

Momani, P.E.N.M., 2009. Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. Am. J. Environ. Sci., 5: 599-604. DOI: 10.3844/ajessp.2009.599.604

Quatieri, T.F., 2002. Discrete-Time Speech Signal Processing. 1st Edn., Pearson Education India, Delhi, ISBN: 9788177587463, pp: 802.

Rabiner, L., M. Cheng, A. Rosenberg and C. McGonegal, 1976. A comparative performance study of several pitch detection algorithms. IEEE Trans. Acoustics, Speech Signal Process., 24: 399-418. DOI: 10.1109/TASSP.1976.1162846

Samad, S.A., A. Hussain and K.F. Low, 2000. Pitch detection of speech signals using the cross-correlation technique. IEEE Proc., 1: 283-286. DOI: 10.1109/TENCON.2000.893673

Schoentgen, J., 2003. Decomposition of vocal cycle length perturbations into vocal jitter and vocal microtremor and comparison of their size in normophonic speakers. J. Voice, 17: 114-125. DOI: 10.1016/S0892-1997(03)00014-6

Swee, T.T., S.H.S. Salleh and M.R. Jamaludin, 2010. Speech pitch detection using Short-Time Energy. Proceedings of the International Conference on Computer and Communication Engineering, May 11-12, IEEE Xplore Press, Kuala Lumpur, pp: 1-6. DOI: 10.1109/ICCCE.2010.5556836

Tabrikian, J., S. Dubnov and Y. Dickalov, 2004. Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model. IEEE Trans. Speech Audio Process., 12: 76-87. DOI: 10.1109/TSA.2003.819950

Timmermans, B., M.S.D. Bodt, F.L. Wuyts, A. Boutewijns and G. Clement *et al*., 2002. Poor voice quality in future elite vocal performers and professional voice users. J. Voice, 16: 372-382. DOI: 10.1016/S0892-1997(02)00108-X