International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 05, May-2015

Robust Pitch Detection Algorithm of Pathological Speech Based on ACF and AMDF

Belgacem Haythem
OS2E Laboratory- The Sciences Faculty of Tunis
Campus of EL MANAR 1060
Tunis Tunisia

Cherif Adnane
OS2E Laboratory- The Sciences Faculty of Tunis
Campus of EL MANAR 1060
Tunis Tunisia

***Abstract*— In this paper, we present a new robust algorithm involving the pre-processing and the extraction of pitch pattern. This method combine the autocorrelation function (ACF) and the AMDF (Average Magnitude Difference Function) to take the advantage of their complementary nature. The falling trends are eliminated by an alignment technique. ACF and AMDF are multiplied and added for several pass band filters to obtain a correct pitch. We present the implementation and the basic experiments and discussions for the proposed algorithm.**

Keywords—ACF; AMDF; Center Clippin; Infinite Pack Clipping; Pitch

I. INTRODUCTION

To make a pathological speech transformation to improve the intelligibility and slow its flow for better comprehension, we use the TD-PSOLA (Time-Domain Pitch Synchronous Overlap and Add) technique which can easily change the flow of speech and the pitch contour. This technique is, chosen for its low computational cost and his simplicity.

The quality of the synthesis of the TD-PSOLA algorithm depends greatly on the chosen pitch detection algorithm. The success of this technique requires a very precise marking of fundamental periods (pitch) on these units to be concatenated. [1] TD-PSOLA is efficient only when the location of the pitch marks, which decompose the signal into overlapping windows synchronized to the fundamental frequency, is very accurate. Thus a decision criterion for classification and labelling into voiced and unvoiced frames was automated by using an artificial neural network (ANN).

The pathological sounds are generally due to changes in geometrical and mechanical properties of the vocal cords and mostly in asymmetric relative to the mid-sagittal plane. Taking into account the complexity of the speech signal in general, and pathological speech in particular, we should choose simple and effective techniques. Pitch marking is a time consuming and error prone task, which has been tackled by several approaches. Present classic pitch detection techniques are more robust, but are unable to process all types of sounds.

In this paper, we propose to implement a new robust algorithm for pitch detection of normal and pathological sounds.

Firstly we present the form and shortcoming of the ACF and AMDF. Then we describe our proposed algorithm in detail with results and discussions.

II. BACKGROUND : REVIEW OF AMDF AND ACF PITCH DETECTION ALGORITHM

Our algorithm is essentially based on classical techniques ACF and AMDF; we present first their basic principles.

A. Autocorrelation Function (ACF)

Autocorrelation is the cross-correlation of a signal with itself. Informally, it is the similarity between observations as a function of the time separation between them. It is a mathematical tool for finding repeating patterns, such as identifying the fundamental frequency in a signal implied by its harmonic frequencies. ACF is often used for analysing functions or series of values, such as time domain signals. [2]

Autocorrelation function of signal _(k) is defined by 1

$$\phi_n(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(m).x(m+k). \qquad (1)$$

We can conclude that the autocorrelation of a periodic signal presents maximums for the moments -2T; -T; 0; +T; +2T. These maxima are called the picks.

B. Average Magnitude Difference Function (AMDF)

The concept of AMDF is very close to ACF, except that it estimates the distance insted of similariy betwen a frame x(m), and its delayed version.[3] AMDF is defined by the following formula 2

$$D(k) = \frac{1}{N}\sum_{n=0}^{N-1} |x(m) - x(m+k)| \qquad for \quad k = 0, ...K. \qquad (2)$$

It is clear that if the wave was perfectly periodic, we should observe D(i) = 0; i = 0; 1::: the practice shows that the estimation of the pitch by the search for a low result in D(k) is fairly easy despite the non-stationarity. This method can be applied to signal slices that contain at least one period of the pitch. The detection of the fundamental is making by detecting zero values of D(k) in the corresponding voiced at T0. [3]

III. THE PROPOSED PITCH DETECTION ALGORITHM

The proposed algorithm for the detection of pathological pitch sounds, combines ACF and AMDF techniques. A post processing is performed on the frames to be analyzed with an appropriate filter by a filter bank, an alignment with the technique of center clipping (CC)before applying the AMDF, center clipping and infinite peak clipping (IPC) before ACF.
Pre-treatment downstream is applied at the end, to select the pitch from the candidates found. "Fig. 1" Shows the block diagram of the proposed pitch detection algorithm and its working procedure is as follows:

1) Filtering the frame: At the input of the pathologic frame passed through five band pass filters
2) Processing to the output of each filter:
a) AMDF:
_ the center-clipping is applied to the frame,
_ the technique of the AMDF is called.
_ Mirroring is applied
b) ACF:
_ the center-clipping is applied to the frame
_ the infinite clipping package is added
_ the technique of the ACF is called
c) Combining AMDF and ACF:
_ the mirror results AMDF and ACF is multiplied
3) Combining all filters:
The products of the 5 filters are combined by addition to get the best candidates finalists.
4) Candidate selection:
The candidate peaks are sorted according to the peak amplitude.

A. *Pre-processing*

The pathological sound is a signal very rich in harmonic components. Thus, the signal may contain 30-40 harmonic components. As the first formant is generally between 300-800 Hz fundamental component is not often the strongest, the trajectory of the formants (F1, F2, F3) is not linear.(see Fig. 2).
For a pathological sound, the F0 value varies between 70 Hz and 500 Hz;[4], so the frequency components above 500 Hz are useless. Thus, a low pass filter with bandwidth of frequencies slightly above 500 Hz would be needed to eliminate the unwanted harmonies. A filtered using a bank of five band pass filters (50- 200Hz, 150-300Hz, 250-400Hz, 350-500Hz, and 450-550 Hz) is placed at the input.
To reduce the effects of formants, an alignment with the technique of center clipping is used. The relationship between input x(n) and output y(n) is given by 3:

$$y(n) = clc[x(n)] = \begin{cases} x(n) - C_L, & if \quad x(n) \geq C_L \\ 0, & if \quad |x(n)| < C_L \\ x(n) + C_L, & if \quad x(n) \leq -C_L \end{cases} \quad (3)$$

Where, CL is the clipping threshold. CL is generally about 30% of the maximum amplitude of the signal [5]. In practice CL should be as high as possible. The Equation 4 is used to determine CL, with A and B, are respectively the values of the first peak of the 150 first and the 150 last samples of the frame (the length of the frame is 450 points)

$$CL = 0{,}66 * \min(A,B). \quad (4)$$

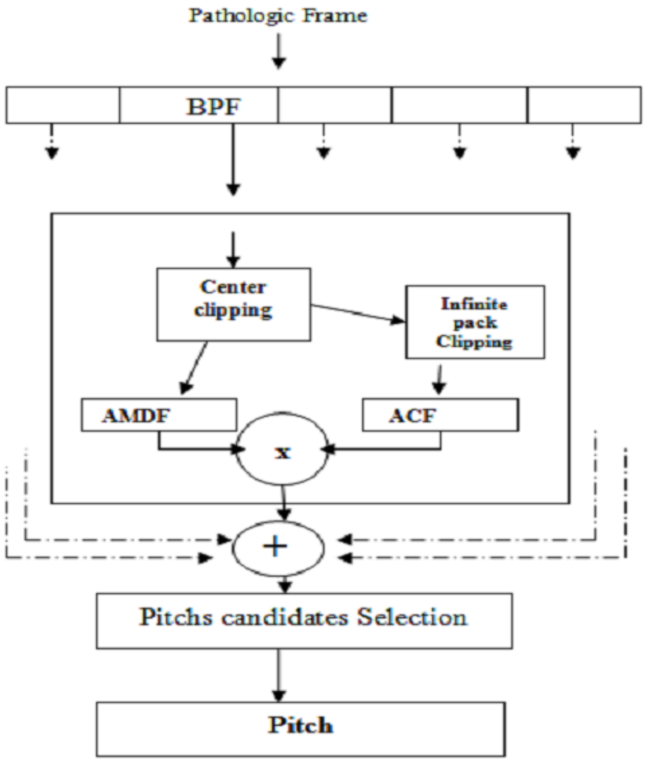

Fig. 1. Bloc diagram of the proposed method of pitch detection based on AMDF and ACF.

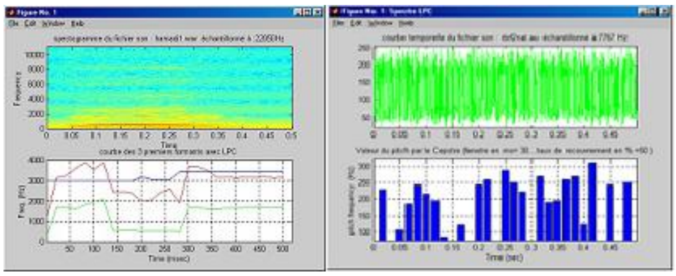

Fig. 2. (Right)Trajectory of F1, F2, F3 for a pathologic sound (left) Pitch variation

Once the center-clipping is applied to the frame, the technique of the AMDF is called. For the short-term autocorrelation, we add another nonlinear treatment which is the infinite clipping package given by equation 5 [6]

$$y(n) = sgn[x(n)] = \begin{cases} 1 & if \quad x(n) \geq C_L \\ 0 & if \quad |x(n)| < C_L \\ -1 & if \quad x(n) \leq -C_L \end{cases} \quad (5)$$

B. Post processing

We make the combination of two chosen methods: AMDF and ACF, AMDF supplies minimum while, the ACF supplies peaks. A mirror effect is applied to the AMDF. We combine the mirror results AMDF and ACF by multiplication and we obtain five products, one for each filter. This multiplication is applied to reduce the number of undesirable candidates. These products are combined for all filters by addition to get the best candidates finalists. Candidate selection is realized by the search of the peak (local maximum) P_0,P_1, …P_{k-1} from each frame. The candidate peaks are sorted according to the peak amplitude. This new method provided a better estimate of the pitch of speech signal. (See "Fig. 4")

TABLE I. COMPARISON OF %PER OF FIVE PITCH DETECTION METHODS

Methods	Normal voices		pathologic voices	
	Male	*Female*	*Male*	*Female*
ACF	14,74	18,27	25,56	27,13
ACF with Clipping, IPC	10,60	11,12	19,03	19,98
AMDF	9,67	11,55	20,33	23,36
AMDF with Clipping	6,25	7,40	17,42	18,26
Our Proposed algorithm	5,20	6,81	8,76	9,05

IV. RESULTS AND DISCUSSIONS

180 examples of sound from the OSEE database, 90 files of pathological sounds and 90 normal files (50% men voices and 50% women voices) are used for this experiment. The sounds are sampled at 16 KHz, 16 bit. The pitch detection results are expressed as a percentage of Pitch Error Rate (% PER) and the Global Pitch Error (%GPE). If is estimated value of _1ms reference, is then considered as warring. The % PER is calculated for the male and female patients as follows:

$$PitchErrorRate = \frac{countoferrorpitch}{countsofallpitch}\% \quad (6)$$

The %GPE of a method is the average of the 4 % PER of each family of sounds: normal male, normal female, pathological male, pathological female.

$$GPE = \frac{PER_{norM} + PER_{norF} + PER_{pathoM} + PER_{pathoF}}{4} \quad (7)$$

The Table. 1. shows the percentages of PER for our proposed method and 4 selected methods of pitch detection., we can noticed that the percentage passes of 14,74% to 5,20% for a male sounds, and from 18,27% to 7,40% for a female sounds by applying our proposed method when compared to the classic ACF.

Also for the pathological sounds the percentage passes from 25,56% to 8,76% for male sounds, and from 27,13% to 9.05% for female sounds by applying our proposed method. fig. 3. also shows the performance of different methods of pitch detection in% PER.

The Table. 2. shows the percentages of GPR for our proposed method and 4 pitch detection methods, we noticed that the percentage goes from 21,42% for the conventional ACF method, to 7,45% for our proposed method. To examine the robustness of our algorithm, we test our proposed method and 4 classic selected methods of pitch detection in different noisy environments.

TABLE II. COMPARISON OF THE % GPR OF 5PITCH DETECTION METHODS

Methods	%GPE
ACF	21,42
ACF with Clipping, IPC	15,18
AMDF	16,22
AMDF with Clipping	12,33
Our Proposed algorithm	7,45

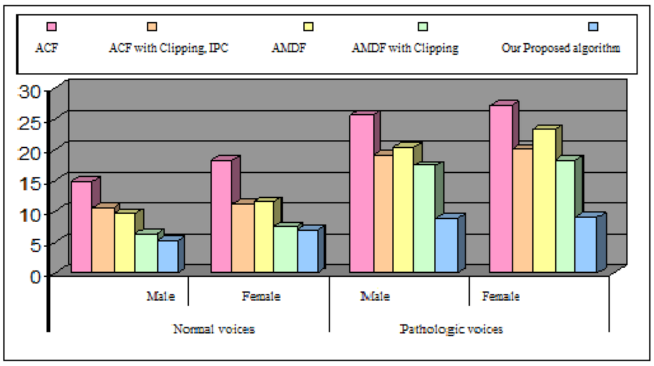

Fig. 3. Performance of different methods of pitch detection on %PER

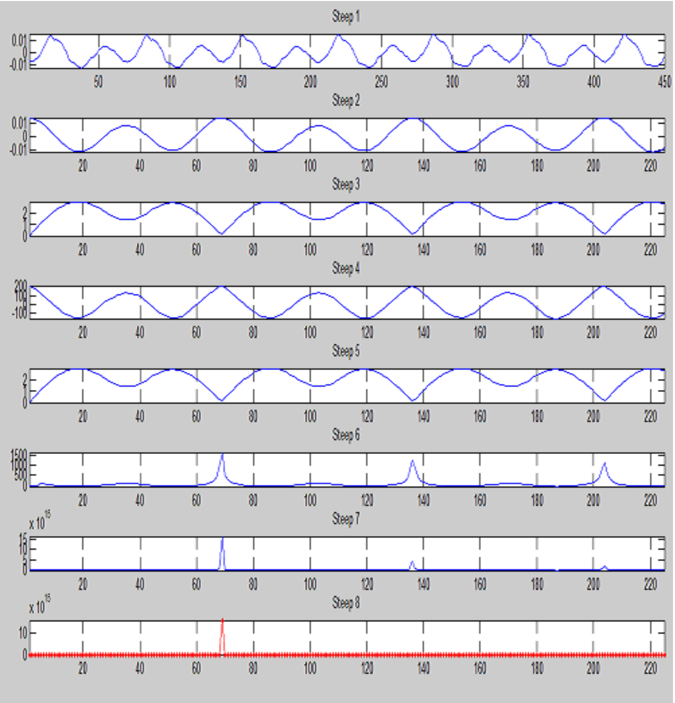

Fig. 4. Pitch Detection steeps with our proposed method. (steep 1) original frame; (steep 2) ACF method; (steep 3) AMDF method; (steep 4) ACF with CC and IPC ;(steep 5) AMDF with CC; (steep 6) multiplication of ACF and AMDF; (steep 7) addition result of all the filters; (steep 8) the final pitch selected from pitch candidates.

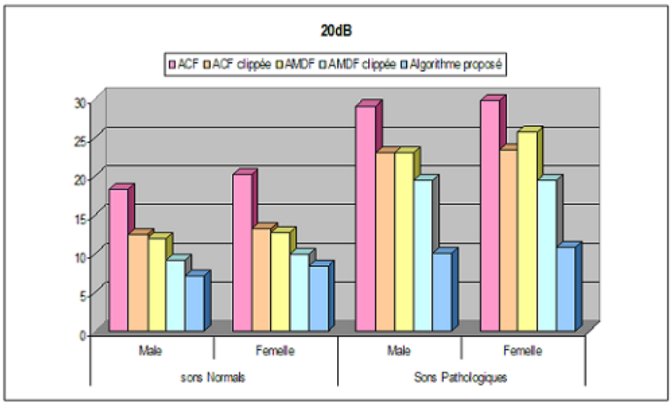

Fig. 5. The % PER of 5 Pitch detection methods for SNR= 20dB.

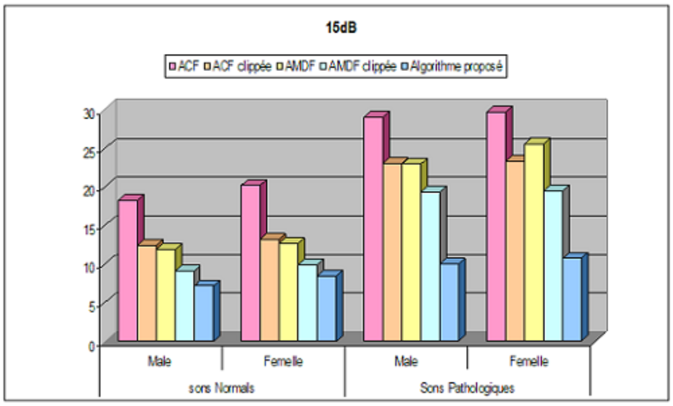

Fig. 6. The % PER of 5 Pitch detection methods for SNR= 15dB.

After adding a noise with signal-to-noise ratios equal to SNR = 20 dB, 15 dB, 10 dB and 5 db.

The" Fig. 5"," Fig. 6", "Fig. 7", "Fig. 8" shows the performance of different methods of pitch detection in% PER. For SNR = 20 db, SNR = 15 db, SNR = 10 db, SNR = 5db.

We notice that for SNR = 10 db, the percentage passes from 26, 47% to 12, 63% for male sounds using the proposed method when compared to the classic ACF. For pathological sounds the percentage passes from 35, 78% to 15, 66%.

The Table 3 includes the percentages of GPE for 5 pitch detection methods in different experimental conditions with different SNR. So we can see that the proposed method has the lowest percentage. This percentage passes from 29,66% to 12,86% using the proposed method when compared to the classic ACF

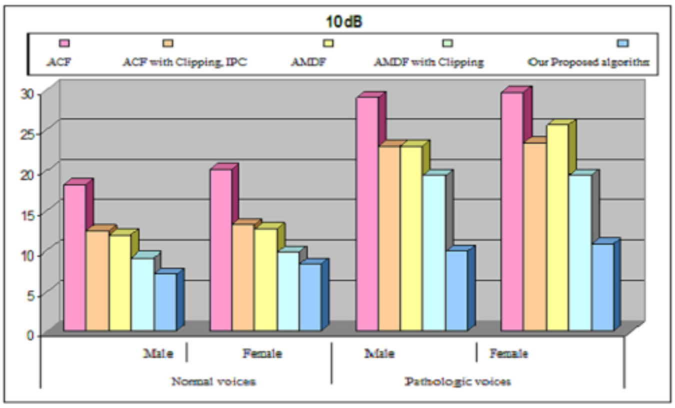

Fig. 7. The % PER of 5 Pitch detection methods for SNR= 10dB

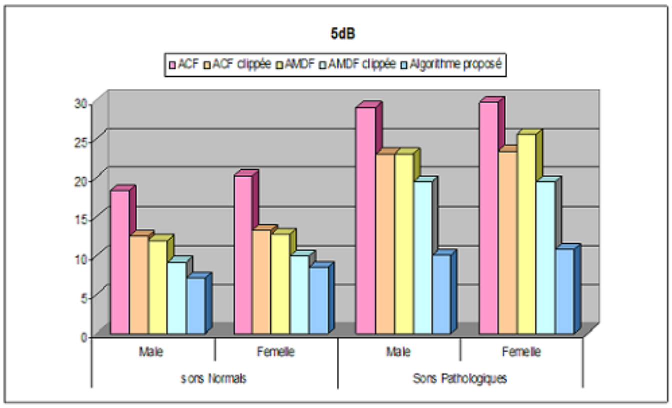

Fig. 8. The % PER of 5 Pitch detection methods for SNR= 5dB

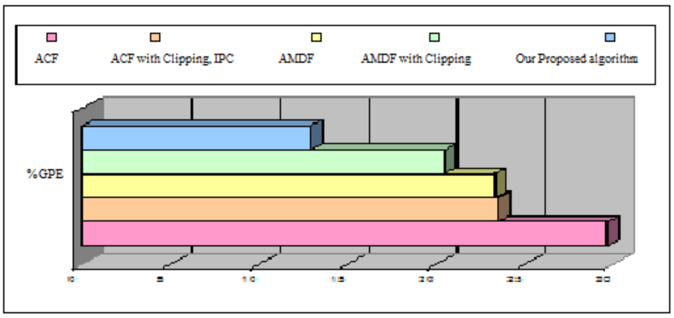

Fig. 9. The % GPR of 5 Pitch detection methods

TABLE III. COMPARISON OF THE % GPR OF 5 PITCH DETECTION METHODS IN NOISE

Methods	%GPE
ACF	29,66
ACF with Clipping, IPC	23,49
AMDF	23,30
AMDF with Clipping	20,46
Our Proposed algorithm	12,86

V. CONCLUSION

Determining the fundamental period of the pathological speech signal by the habitual methods (AMDF, ACF, ACF with clipping and infinite clipping pack, with clipping AMDF) miss performance for pitch detection especially for pathological sounds.

The ACF method despite its simplicity of use these results in real time presents problems in cases where the peaks due to the response of the vocal tract are larger than those due to the periodicity of the excitation speech, is consequently dependent on the stationarity of the speech signal, a condition which is not always true for pathological sounds.

The AMDF method, it does not rely on the stationarity of the signal as it reduces the ambiguity between the peaks and harmonics of the fundamental. This method gives better results with large window size; the non-stationarity of pathological speech signal prevents us from answering this requirement. In this paper, we have presented a robust pitch detection algorithm of pathological pitch sounds, combines techniques ACF and AMDF with alignment post processing and a selection of candidates. Its efficiency and effectiveness has been validated by several experiments.

This new method appears robust to irregular pathological sounds, it can outperform other methods considering the tradeoffs between computing time and precision.

REFERENCES

[1] S.A Toma, G.I. Tarsa, ; E. Oancea, ; D. Munteanu, “A TD-PSOLA based method for speech synthesis and compression, ” 8th International Conference on Communications (COMM) Bucharest, 10-12 June 2010, pp. 123-126.

[2] H Zhao, W Gan “A New Pitch Estimation Method Based on AMDF, ” Journal of multimedia, October 2013 pp. 618-621, vol.8, NO. 5 .

[3] E. Moulines, F. Emerard, L. Larreur,“A real-time French text-to-speech system generating high-quality synthetic speech,” in ICASSP-90, International Conference , 3 6 April,1990, pp. 309-312, vol.1 .

[4] A. Cherif “Pitch detection and formant extraction of Arabic speech processing ”Journal of applied acoustics, January 2001.

[5] S.S. Nimbhore, G.D. Ramteke, R.J. Ramteke “Pitch estimation of Marathi spoken numbers in various speech signals ,” International Conference on Communications and Signal Processing (ICCSP), 2013.

[6] M.M. Sondhi, “New methods of pitch extraction, ” IEEE Trans.Audio Electroacoust., vol. AU-16, pp. 262-266, June 1968

[7] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, C. A. McGonegal. “A comparative performance study of several pitch detection algorithms,” IEEE Transactions on Audio, Signal, and Speech Processing, pp 399 417, 1976

[8] L. Tan, M. Karnjanadecha.“Pitch Detection Algorithm: Autocorrelation Method and AMDF”roceedings of the 3rd International Symposium on Communications and Information Technology pp.551 556.september 2003

[9] C. Shahnaz, P. Zhu“A robust pitch estimation algorithm in noise ”ICASSP2007 April 16 20, 2007. Hawaii, USA. pp.551 556.september 2007

[10] C. Manfredi, M. DAniello, P. Bruscaglioni, A. Ismaelli “A comparative analysis of fundamental frequency estimation methods with application to pathological voices”Medical Engineering and Physics pp.135 147.2000

[11] R. Ritchings, M.A. Mcgillion, C.J. Moore“Pathological voice quality assessment using artificial neural network ” Medical Engineering Physics pp 561-564, ELSEVIER, 2002.

[12] H. BELGACEM, A. Cherif“Automatic determination of pathological voice transformation coefficients for TDPSOLA using neural network” International Multi-Conference on Systems, Signals and Devices SSD pp.135 147 .1569387415 SSD11March 2011 Sousse Tunisia