Analysis and Parameter Extraction of P Wave Using Correlation Method

Mohamed Ben Messaoud, Besma Khelil, and Abdennaceur Kachouri
Laboratory of Electronic and Information Technology, National School of Engineers of Sfax, Tunisia

Abstract: *The objective of this paper is to develop an efficient P wave detection algorithm based on the morphology characteristics of arrhythmias using correlation and regression in ECG signal. Subjects for experiments included normal subjects, patients with atrial fibrillation, ventricular tachycardia, and patients with change of the artifactuale amplitude. After the step of the detection of R peak using the pan- tompkins algorithm, the correlation and regression were utilized to calculate the similarity factors between a studied P wave and the reference one. The correlation coefficient can indicate the kind of arrhythmia diseases. The algorithm was tested using MIT-BIH arrhythmia database where every P wave was classified. The results are presented in terms of correlation coefficient. Then some parameters have been extracted in order to classify the arrhythmias. The correlation coefficient results of the system are 1,0.07 and –0.92 for normal beats, atrial fibrillation and change of the artifactuale amplitude, respectively. The extracted parameters are closely similar to the expert values given by the cardiologist. The results reveal that the system is accurate and efficient to detect and classify arrhythmias resulted from atrial fibrillation or change of the artifactuale amplitude.*

Keywords: *ECG, P wave detection, correlation, regression, diagnosis.*

Received April 18, 2007; accepted October 23, 2007

1. Introduction

The analysis of the electrocardiogram signal is the most readily available method for diagnosing cardiac arrhythmias. In electrocardiograms such arrhythmias, ECGs manifest themselves as deformations in the observed waveform. Such deformations, as associated with a diagnosed arrhythmia, occur with a consistency and morphological similarity that they may be looked upon as a waveform pattern in the temporal domain.

In order to interpret the ECG and use it to diagnose abnormalities, it is important to know the normal characteristics of the ECG, and understand the mechanisms underlying the generation of each segment of the ECG. Figure 1 shows the various fiducial points in the ECG, and typical of the various intervals measured values from the ECG.

The *P* wave is caused by atrial depolarization. The duration is normally not greater than 120ms. The normal shape of the *P* wave does not include any notches or peaks. It can be positive, negative, or biphasic in the remaining leads. An absent *P* wave in the ECG may signify sinoatrial block [9].

The purpose of this study is to develop a method to distinguish healthy and abnormal subjects using the correlation coefficients of ECG waveforms. Although many detection algorithms for ECG signal have been developed to detect QRS complex, but there are only a fewer publications that describe algorithms to detect the *P* wave detection [9, 17, 11, 15, 6].

The correlation method was used to detect QRS complex [4]. Moreover, the detection of arrhythmia does not rely on *P* wave since it is hard to detect due to noisily signal.

The objective of this paper is to describe an automatic algorithm to detect *P* wave by using coefficient correlation and to extract its parameters in order to make the prevention and the precocious diagnosis of the anomalies of the *P* wave. The Pan-Tompkins algorithm is applied to determine the end of *P* wave. The correlation factor is calculated from the reference *P* wave and the signal under test. The delayed *P* wave is performed to calculate the sign and the maximum correlation factor. These parameters indicate some information about the normal or abnormal ECG.

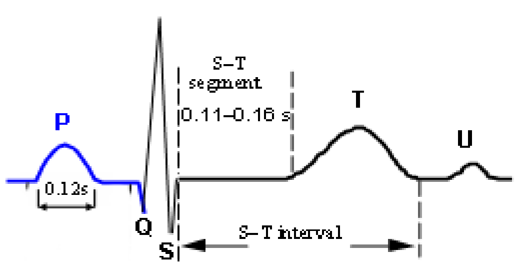


Figure 1. The most important parameters of an ECG signal.

2. Methods

The detection algorithms of *P* wave are generally designed according to two steps. The first is based on a preliminary localisation of the QRS complex and consists in seeking, in a window preceding the QRS, a trace of the *P* wave. The second is based on the detection of the *P* wave by the method of correlation.

In this application, the signals from MIT-BIH database with annotations from specialists are analysed. The methods were developed under Matlab software (V6.5). The description of the *P* wave detection algorithm is shown in Figure 2.

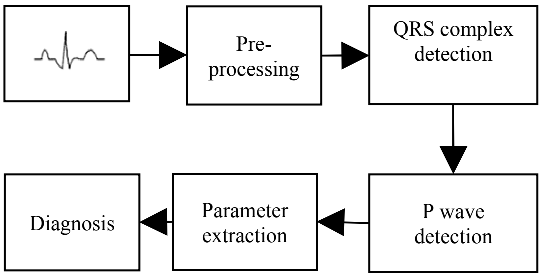


Figure 2. Description of the *P* wave detection algorithm.

2.1. Pre-Processing of the ECG

The ECG signal gotten at the time of the registration is contaminated generally by different sources of noises [12] that can disrupt the phase and the amplitude characteristics of the useful signal, from where the necessity of a good filtering [2].

The ECG signals are filtered already by an anti-withdrawal filter to eliminate the undesirable frequencies (parasitic, noises). But to have some good results, we did a filtering in order to eliminate the residual noises. Wavelet based filters cabe used [3].

We apply a band pass filter [2] to eliminate the noises caused by the breathing, the movements of muscles and the baseline. This filter is constituted by the combination of a passes low filter and another passes loud. The two are based on a sampling rate of 360Hz [8].

2.2. QRS Complex Detection

In general, the normal ECG rhythm means that there is a regular rhythm and waveform. However, the ECG rhythm of the patient with arrhythmia will not be regular in certain QRS complexes. We use the different characteristics that arrhythmias exhibit to detect the abnormal ECG waveform. We first need to find the location of every QRS complex.

The locations of QRS complex have the maximum variation in the slopes. This property was used to detect the location of QRS complex. The method of "Pan and Tompkins" was adopted to detect the QRS complex [14, 10]. The property of variation in slopes and an adaptive threshold were applied to detect the R point.

The "Pan and Tompkins" QRS detection algorithm identifies the QRS complexes based upon digital analysis of slope, amplitude, and width of the ECG data. The algorithm implements a special digital band pass filter. It can reduce false detection caused by the various types of interference present in the ECG signal. The algorithm automatically adjusts the thresholds and parameters periodically to adapt the changes in QRS morphology and heart rate. Details of the algorithm can be found in [14]. In summary, it consists of the following processing steps as shown in Figure 3:

- Band-pass filtering.
- Differentiation.
- Squaring.
- Moving window integration.
- Thresholds adjustment.

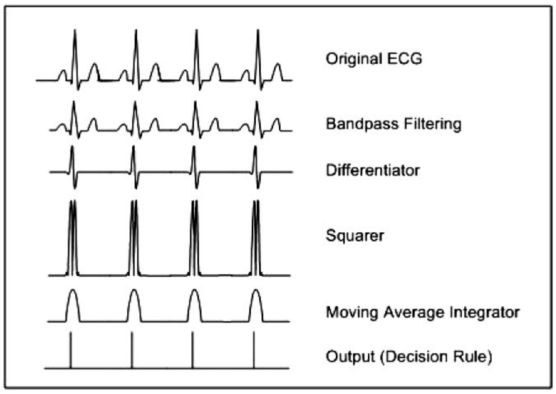


Figure 3. The processing steps of QRS complex detection.

The *P* wave is a rounded peak occurred before the QRS complex. Therefore, the *P* wave can be found based on the location of the QRS complex.

2.3. P Wave Detection

A representative *P* wave template was generated by coherent averaging of the respective signal part of all regular heartbeats. Subsequently, all regular and abnormal heartbeats were compared to this P-wave template using linear regression analysis [16].

P wave detection is carried out based on the correlation coefficient approach. A window is positioned to the *P* wave for the matching of the constituent elements. The window only spans the duration of the *P* wave rather the whole cardiac cycle.

The method consists in checking the correlation between two different cardiac signals for example x and y, of which the goal to look for y in the signal what is the small tip of signal that looks the more like the x signal. For it, x is called mask, that represents in our case the *P* wave of the ECG signal, and one looks for the place most similar to the mask x. The $x(k)$ signal that resembles the more to $y(k)$ is the one that has the maximal correlation coefficient.

2.3.1. Correlation

The correlation is a statistical technique which can show whether and how strongly pairs of variables are related [13]. Like all statistical techniques, correlation is only appropriate for certain kinds of data. Correlation is applied for data in which numbers are meaningful, usually quantities of some sort.

Most statisticians say that you cannot use correlations with rating scales; because the mathematics of the technique assumes the differences between sampling times are equal. Nevertheless, many survey researchers use correlations with rating scales, because the results usually reflect the real world [16]. When we use quantities, correlations provide precise measurements. When we work with rating scales, correlations provide general indications. The main result of a correlation is called the correlation coefficient (or r).

2.3.2. Correlation Coefficient

The correlation coefficient (r) between two random variables x and y was defined as equation 1 [4]:

$$r = \frac{\sum_{n=0}^{N-1}\left[x(n)-\bar{x}\right]\left[y(n-k)-\bar{y}\right]}{\sqrt{\sum_{n=0}^{N-1}\left[x(n)-\bar{x}\right]^2\sum_{n=0}^{N-1}\left[y(n-k)-\bar{y}\right]^2}} \qquad (1)$$

where r is the correlation coefficient, N the number of template points, $x(n)$ the template points, $y(n)$ the signal points under analysis, $\bar{x}$ the average of the template points, $\bar{y}$ the average of the signal points, and k is the time index of the signal $y(n)$ at which the template is placed.

The interval variation of the correlation coefficient is [-1, +1]. When the variables are closely related, r is equal to +1 or -1. If r is close to 0, it means there is no relationship between the variables. If r is positive, it means that as one variable gets larger the other gets larger. If r is negative it means that as one gets larger, the other gets smaller (often called an “inverse” correlation).

2.3.3. Regression

Simple regression is used to examine the relationship between one dependent and one independent variable [5]. After performing an analysis, the regression statistics can be used to predict the dependence variable when the independent variable is known. Regression goes beyond correlation by adding prediction capabilities.

The regression line (equation 2) is a plot of the expected value of the dependence variable for all values of the independent variable.

$$y = a\,x + b \qquad (2)$$

Technically, it is the line that “minimizes the squared residuals”. The regression line is the one that best fits the data on a scatter plot as represented as shown in Figure 4.

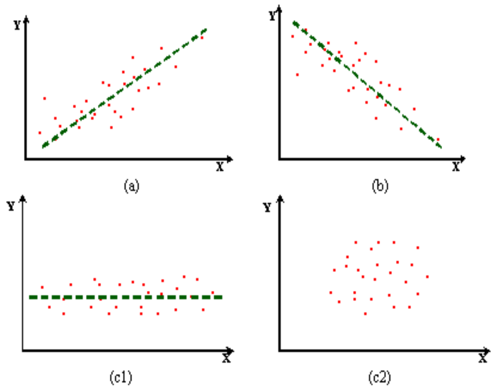


Figure 4. Correlation and regression: (a) Positive correlation, (b) Negative correlation, (c1, c2) Absence of correlation.

Using the regression equation 2, the dependence variable may be predicted from the independent variable. The slope of the regression line *(a)* is defined as the rise divided by the run. The y intercept (b) is the point on the y axis where the regression line would intercept the y axis. The slope and y intercept are incorporated into the regression equation. The intercept is usually called the constant, and the slope is referred to as the coefficient.

2.4. The Parameters

The parameters used to identify the P wave can be classified in several categories [10]. We distinguish notably those described in temporal domain and in the frequency domain. The first are more or less directly accessible to an assessment on the tracing of the ECG, the second require the use of means more complex calculus.

2.4.1. Temporal Parameters

- The length: an elongation of the width of the P wave can indicates to the physicians the presence of electric conduction unrest to the level of the auricles. A width superior to 120 ms (we also find a doorstep to 130ms) identifies an illness to the level of the auricles, either to the level of the electric conduction.
- The shape: the shapes of the P waves were classified on 4 classes: the normal shapes (symmetrical), the shapes with slow ascending slopes, the shapes with slow downward slopes, finally the bi-phasic shapes.

2.4.2. Frequencies Parameters

The use of a time-frequency or time-scale method [1] permits to keep a temporal localization notion and to avoid in particular that the content frequencies of the complex QRS doesn't conceal the one of the P wave. The two main frequencies parameters are the energy and the entropy. The flow chart of the detection algorithm is shown in Figure 5.

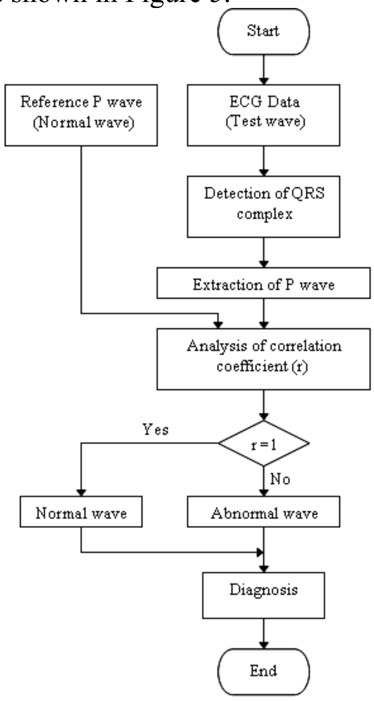


Figure 5. Flow chart of the *P* wave detection algorithm.

3. Simulation Results

3.1. DataBase

The ECG signals under test of the transformations are taken from the website of MIT-BIH Arrhythmia database [7]. The sampling frequency of the data is 360Hz. The references of the signals used and their descriptions appear as shown in Table 1. Table 2 contains the features of a normal ECG, and the different anomalies that we will study.

Table 1. The studied MIT recording.

Reference of the Recording	Description
MIT100	Normal rhythm
MIT202	Atrial fibrillation
MIT203	Ventricular tachycardia
MIT214	Change of the artifactuale amplitude

Table 2. Characteristic of the different cardiac anomalies.

Anomalies	Feature
Normal case	- *P* wave (amplitude of 0.1 to 0.3mV, length < 0.12s) - Length of QRS $\leq$ 0.12s
Atrial fibrillation	- Absence of the P wave - Irregularity of the rhythm
Ventricular tachycardia	- Large QRS complex (> 0.12s), anarchical
Change of the artifactuale amplitude	- Large QRS - Negative P wave

Figure 6 presents the recordings ECG of the MIT BIH database.

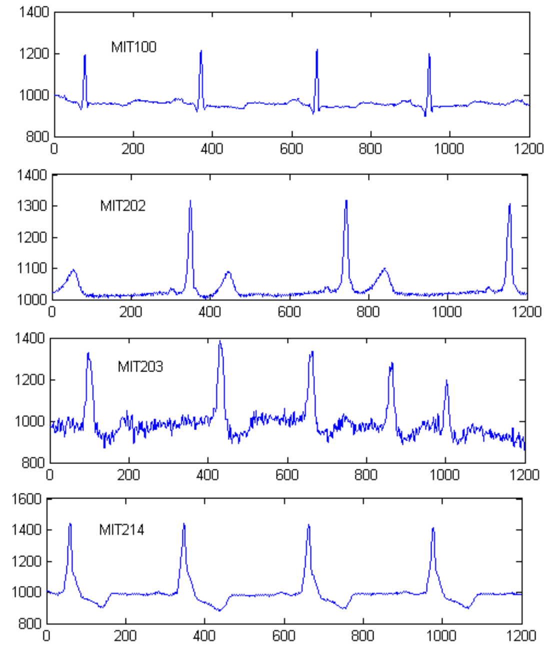
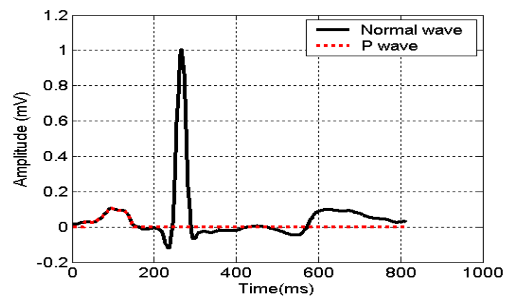


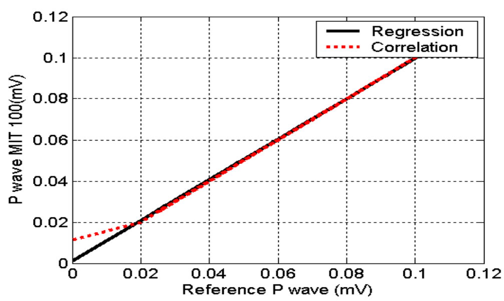
Figure 6. The studied ECG from the MIT BIH database.

3.2. Simulation

For each recording, we carry out the superposition between a reference *P* wave and a cardiac cycle. Let us take the case of ECG signal refereed MIT100 as shown in Figure 7(a), the correlation curve between the two signals and the regression line was shown in Figure 7(b).



(a) ECG signal and normal P wave.



(b) Correlation between the 2 signals and the regression line.

Figure 7. Recording MIT 100.

The crucial problem in *P* wave duration measurement consists in the definition of the fiducially points. The starting and the ending of the *P* wave are detected either manually or automatically. In some cases they are defined as the points corresponding respectively to the first (onset) and the last (offset) deflection from the baseline [18].

In this paper, the variation of correlation coefficient according to delta, which represents the temporal shift, was used to identify the *P* wave duration (width). The shift varies between the moment t_0 which indicates the onset of QRS complex and the beginning of ECG signal (PR interval) as shown in Figure 8.

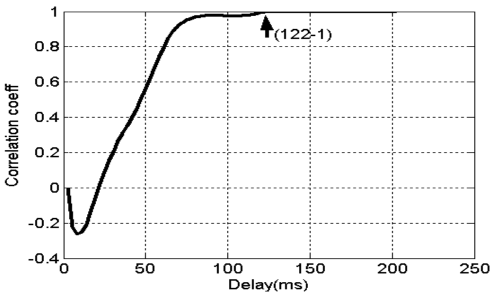
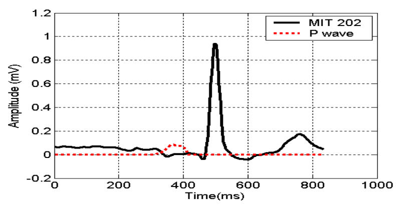


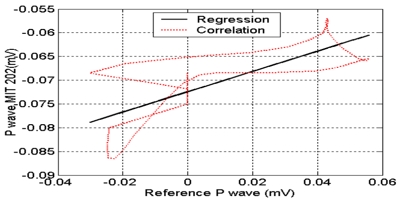
Figure 8. Variation of correlation cœfficient according to the shift (MIT100).

The value of delta, which corresponds to the maximum value of the correlation coefficient, multiplied by Te (scaling time) indicates the width of the *P* wave. The amplitude of *P* wave is the maximum value of this wave.

Figure 9 presents the correlation between signal ECG, MIT202 and the normal *P* wave. The variation of correlation coefficient according to the shift is shown in Figure 10.



(a) ECG signal and normal P wave.



(b) Correlation between the two signals and the regression line.

Figure 9. Recording MIT202.

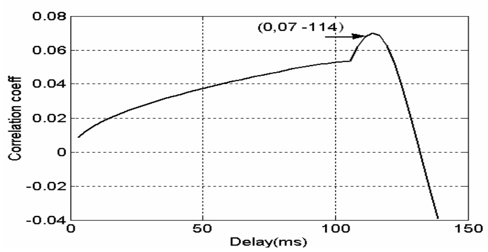
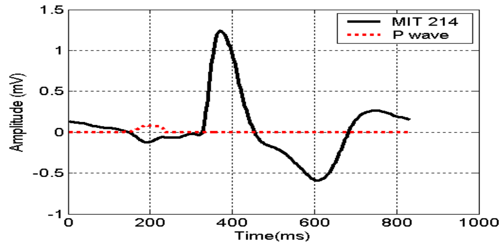
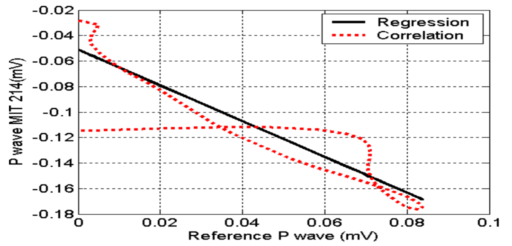


Figure 10. Variation of correlation cœfficient according to the shift (MIT202).

In Figure 11, the correlation between signals ECG, MIT214 and the normal *P* wave are presented. The curve of variation of the correlation coefficient according to the shift is given by Figure 12.



(a) ECG signal and normal P wave.



(b) Correlation between the two signals and the regression line.

Figure 11. MIT214.

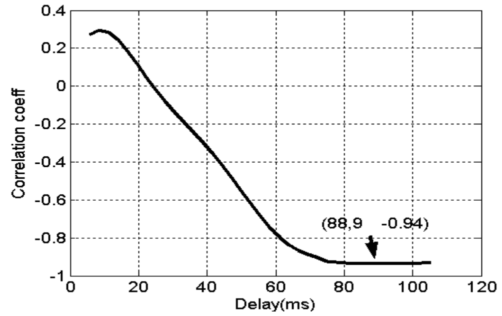


Figure 12. Variation of correlation coefficient according to the shift (MIT214).

From the previous curves, the extraction of the following parameters relative of each signal is deduced: correlation coefficient, amplitude and the width of *P* wave. The Table 3 summarizes the simulation results for the different signals.

Table 3. Detection of the *P* wave by the correlation method.

ECG Recording	**Correlation Coefficient (R)**	**Amplitude of the *P* Wave (Mv)**	**Width of the P Wave (S)**
MIT100	1	0.11	0.12
MIT202	0.07	0.058	0.11
MIT214	- 0.92	- 0.1	0.089

4. Discussion

For the case of MIT100 signal, the regression line shows that we have a positive correlation between the MIT100 signal and reference *P* wave. The value of *r* increases as far as reaching a maximum equals to 1. The corresponding value of delta is equal to 122ms. Therefore we can conclude that this signal is a normal case of which the length of the *P* wave is equal to 0.12s. In fact that it is very close to the value presented in the Table 4 which contains the parameter values extracted by a cardiologist.

Table 4. Manual extraction of the parameters.

ECG Recording	Amplitude Of The *P* Wave (Mv)	Width Of The *P* Wave (S)
MIT100	0.106	0. 12
MIT202	0.05	0.1
MIT214	-0.1	0.1

In the case of MIT202 signal, the intercept value of regression line is weak and the value of '*r*' is equal to 0.07 as shown in Figure 10, meaning the absence of the *P* wave, which confirms that recording MIT202 is an atrial fibrillation.

For the MIT203 signal, the signal is characterized by an irregular rhythm that changes continually. This ECG is only constituted of the anarchical QRS complex (not of *P* and *T* waves); therefore we are not able to apply our method for this case.

For the case of MIT214 signal, the regression line indicates that we have a negative correlation between the MIT214 signal and reference *P* wave. The value of *r* reaches a minimum negative value for delta equals to 88.9ms as shown in Figure 12. Therefore we note that we have a negative *P* wave. So the correlation coefficient between this signal and the reference *P* wave is equal to –0.94 and it is a negative value mining a negative correlation.

We calculate the relative error (*Er*) between the values given in the Table 3. That represents the parameters automatically extracted with the proposed method and those of the Table 4. that represents the parameters given by a specialist. Relative error is defined as equation 3:

$$E_r = \frac{V_m - V_r}{V_r} \quad (3)$$

where V_m is a measured value (automatic method) and V_r is a real value (manual method), The results are presented in the Table 5.

Table 5. Relative error.

ECG Registrations	Amplitude of the P Wave	Duration of the P Wave
MIT100	4 %	0%
MIT202	14%	10%
MIT214	0%	11 %

The relative error is practically weak; therefore we can consider this method as helpful tool to the diagnosis for the anomalies of the *P* wave especially.

5. Conclusion

The paper presented an automatic method of analysis of the *P* wave of an electrocardiogram. This method is based on the correlation theory; mainly it recognizes the degree of similarity between a reference and arrhythmia *P* wave. The regression line indicates the kind of diseases. In simulation, the cases studied are normal subject, atrial fibrillation and change of the artifactuale amplitude. The results show the ability of the proposed algorithm to the classification purpose. Then for diagnosis, the amplitude and the width of *P* wave are extracted. The comparison with the manually extracted parameters presents a reduced error.

In conclusion, the simulation results, for the detection of the *P* wave using correlation, give acceptable results. The proposed arrhythmia detection algorithm can be a useful method to the clinical diagnosis.

References

[1] Addison P., "Wavelet Transforms and the ECG: A Review," *Physiological Measurement*, vol. 26, no. 5, pp. 155-199, 2005.

[2] Belgacem N., Chikh M., and Reguig F., "Détection et Identification des Arythmies Cardiaques Par Application des Réseaux de Neurones," *in Conférence Internationale sur les Systèmes de Télécommunications d'Electronique Médicale et d'Automatique CISTEMA*, Algeria, pp. 236-239, 2003.

[3] BenMessaoud M., Kachouri A., and Sai R., "Diagnostic of Electrocardiogram ECG by Wavelet Transform Technique," *in Proceedings of the IEEE Symposium on Signal Processing and Information Technology (ISSPIT),* Egypt, pp. 1, 2001.

[4] Chiu C., Lin T., and Liau B., "Using Correlation Coefficient in ECG Waveform for Arrhythmia Detection," *Biomedical Engineering Applications Basis Communications,* vol. 17, no. 3, pp. 147-152. 2005.

[5] Cohen J., Cohen P., West S., and Aiken L., *Applied Multiple Regression/Correlation Analysis fort the Behavioral Sciences*, Lawrence Erlbaum, New York, 1983.

[6] Dubois R., *Application des Nouvelles Méthodes d'Apprentissage à la Détection Précoce d'Anomalies en Electrocardiographie*, Thèse de Doctorat, Paris, 2004.

[7] PhysioNet, http://www.physionet.org/physiobank /database/ht ml /mitdbdir/intro.htm, 2007.

[8] Khelil B., Kachouri A., and Ben Messaoud M., *Nouvelle Tendance Technologiques en Génie Electrique et Informatique,* Springer, New York, 2006.

[9] Le Page R. and Boucher J., *Détection et Analyse de l'Onde P d'Un Electrocardiogramme*, Colloque GRETSI'01, France, 2001.

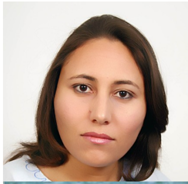
[10] Le Page R., *Détection et Analyse de l'onde P d'Un Electrocardiogramme : Application au Dépistage de la Fibrillation Auriculaire.* Thèse de Doctorat, Paris, 2003.

[11] Mahmoodabadi S., Ahmadian A., and Abolhasani M., "ECG Feature

Extraction Using Daubechies Wavelets," *in Proceedings of the International Association of Science and Technology for Development (IASTED)*, Spain, pp. 220-225, 2005.
[12] McNames J., "Optimal Rate Filters for Biomedical Point Processes," *in Proceedings of the IEEE Medical & Biological Engineering* 27^{th} *Annual*, China, pp 5388-5391, 2005.
[13] Narayan S., Kahn A., Brown J., Krummen D., Bhargava V., and Feld G., "Determination of Intra-Atrial Fibrillatory Cycle Length from the Electrocardiogram Using Auto-Correlation Analysis," *International Journal of Bioelectromagnetism Advances in Electrocardiology,* vol. 7, no. 1, pp. 162-165, 2005.
[14] Pan J. and Tompkins W., "A Real Time QRS Detection Algorithm," *IEEE Transactionsons on Biomedical Engineering*, vol. 32, no. 3, pp. 230-236, 1985.
[15] Portet F., "Pilotage D'algorithmes Pour la Reconnaissance en Ligne D'arythmies Cardiaques," Thèse de doctorat, Paris, 2005.
[16] Schreicr G., Kastner P., and Marko W., "An Automatic ECG Processing Algorithm to Identify Patients Prone to Paroxysmal Atrial Fibrillation," *IEEE Computers in Cardlology*, vol. 28, pp. 133-135, 2001.
[17] Tan K., Chan K., and Choi K., "Detection of the QRS Complex, *P* Wave and *T* Wave in Electrocardiogram," *in Proceddings of First International Conference on Advances in Medical Signal and Information,* pp. 41-47, UK, 2000.
[18] Poli S., Barbaro V., Bartolini P., Calcagnini G., and Censi F., "Prediction of Atrial Fibrillation from Surface ECG: Review of Methods and Algorithms," *in Annual* I^{st} *Super Sanità*, vol. 39, no. 2, pp. 195-203, 2003.

Mohamed Ben Messaoud received the engineer degree in electrical engineering from the University of Sfax, Tunisia, in 1981 and the Docteur Ingenieur Diploma in automatic control engineering from the University of Paul Sabatier of Toulouse, France, in 1983.

Besma Khelil received the engineer degree in electric engineering from the National School of engineering of Sfax (ENIS), Tunisia, in 2005, and the MSc degree in electronic from University of Sfax, Tunisia, in 2007.



Abdennaceur Kachouri received the engineering diploma from National School of Engineers of Sfax in 1981, a Master degree in measurement and instrumen-tation from National School of Bordeaux (ENSERB) of France in 1981, and a Doctorate in measurement and instrumentation from NSERB, in 1983.