Biomedical Signal Processing and Control 4 (2009) 309–316

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Analysis of changes in the beat-to-beat P-wave morphology using clustering techniques

Alberto Herreros [a,*], Enrique Baeyens [a], Rolf Johansson [b], Jonas Carlson [b], José R. Perán [a], Bertil Olsson [c]

[a] Department of Systems Engineering and Automatic Control, University of Valladolid, Paseo del Cauce, s/n, E-47011 Valladolid, Spain
[b] Department of Automatic Control, LTH, Lund University, Box 118, SE-221 00 Lund, Sweden
[c] Department of Cardiology, University Hospital, Lund University, SE-221 85 Lund, Sweden

ARTICLE INFO

Article history:
Received 3 November 2008
Received in revised form 10 February 2009
Accepted 17 February 2009
Available online 31 March 2009

Keywords:
P-wave morphology
K-means clusters
ECG wave delineation
Atrial electrical activity

ABSTRACT

Several pathologies related to the atrial electrical activity can be detected in the electrocardiogram P-wave. A protocol for analyzing P-wave morphology changes has been developed in this article. By using this protocol a study on the beat-to-beat P-wave morphology changes of 89 ECG signals is performed. An algorithm based on the embedding space techniques has been used to extract the P-wave information of the ECG. The P-waves obtained in several of these ECGs exhibit significant alternate morphology changes. The morphologies have been classified by using the *K*-means clustering algorithm. The mechanism behind the P-wave morphology change process and its possible pathophysiological importance remains to be clarified.

1. Introduction

The interatrial conduction delay is considered as one of the causes of several atrial pathologies. The normal cardiac rhythm is initiated in the sinus node, and is propagated along the entire atrial myocardium. There are three internodal conduction pathways in the right atrium and multiple potential interatrial paths. Atrial conduction disorders frequently appear in elderly or structural heart disease subjects and can be the origin of the atrial fibrillation, flutter or tachyarrhythmias, see [9].

The analysis of the electrocardiogram (ECG) is a well-known non-invasive technique to detect the electrical heart activity. A normal heart cycle is reflected in the ECG by a P-wave (atrial activity), a QRS complex and a T-wave. Several clinical studies connect certain P-wave properties such as its width and morphology, with anomalies in the electrical atrial conduction and atrial pathology.

The relation between the P-wave morphology and the interatrial delay has been studied in [3,9]. Recently, it has been clearly verified how P-wave morphology depends upon which of three interatrial routes the conduction from right to left atrium has utilized [13]. The analysis of the P-wave obtained from Frank leads and its relation to atrial fibrillation is reported in [20,19,8] and its relation to age in [6]. The width of the P-wave, the time position of the *X*-lead peak and the sign of the *Z*-lead signal value are some of the most frequently used indices to predict atrial anomalies.

The magnitude and width of the P-waves are smaller than those of the QRS complexes and T-waves, and their study requires more sophisticated computation techniques. The ECG baseline oscillations, due to the respiratory cycle, and the high frequency noise have, in some cases, the same magnitude as the P-wave and could hide or disturb it. Most ECG detection algorithms only detect the QRS complexes, see [15,23]. Other algorithms try to obtain all the characteristic points of the ECG, i.e., the onset, peak and end of the P-waves, QRS complexes and T-waves.

Following [21], the P-wave detection methods can be classified into two groups. The first group searches for the P-wave in a localized area outside the QRS-T region. These algorithms first detect the QRS complex and the T-wave and then search for the P-wave outside this region using different techniques. The most important are differential low pass filtering, wave analysis, hidden Markov models and neural networks, see [17,16]. The algorithms in the second group split the atrial and ventricular activity by direct separation methods or by QRS-T cancellation. The most common technique is the subtraction of a QRS-T template and several methods have been used [11]. In order to improve performance in obtaining the P-wave, different approaches are used to split ventricular and atrial sources. The most important are blind source separation and independent component analysis, see [24].

* Corresponding author.
E-mail addresses: peran@eis.uva.es, albher@eis.uva.es (A. Herreros), enrbae@eis.uva.es (E. Baeyens), Rolf.Johansson@control.lth.se (R. Johansson), Jonas.Carlson@kard.lu.se (J. Carlson), Bertil.Olsson@med.lu.se (B. Olsson).

1746-8094/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bspc.2009.02.006

The main goal of this paper is the development of a new methodology to extract the P-wave information of a medium/long duration ECG and its morphology classification. The method consists of three steps. The first one is the P-wave onset and end points detection using a novel phase-space algorithm. The second one consists of extracting, rotating and filtering the P-wave. And the third one is the classification of the P-waves by using clustering techniques. Finally, a morphology study of the P-wave and its beat-to-beat changes is carried out by both graphical methods and using the indices and criteria proposed in [19,8].

The algorithm reported in [12] is used to detect and extract the P-waves from ECG records. This algorithm avoids the above-mentioned difficulties associated to the detection of the P-wave characteristic events and is based on the embedding phase-space of a measured signal, see [14,2,1]. The algorithm transforms the ECG lead time signal in a sequence of points of a new multi-dimensional space. Its performance has been tested using both simulated and true ECG signals.

The extraction, filtering and rotation of the P-waves are instrumental for the morphology classification. The mathematical tool used for classification is the *K*-means clustering algorithm, see [4,10], that allows a choice in the number of the required classification sets. Other classification approaches such as neural nets could be applied, see [5]. However the *K*-means algorithm is faster, easier to implement and does not require a training phase.

The main reason to classify the P-waves is the detection of beat-to-beat changes in their morphology. In this study, the number of cluster has been always two, because we were interested in classifying individuals into two groups: those that have a constant P-wave morphology and those that suffer of beat-to-beat morphology changes. It is interesting to remark that a finer analysis could be accomplished for those individuals suffering of beat-to-beat morphology changes by using more than two clusters.

The developed protocol has been applied to 89 orthogonal ECG signals belonging to both healthy individuals and patients with atrial rhythm disorders. It has been checked that some of them present significative changes in the beat-to-beat P-wave morphology. The pathological reasons for these morphology changes deserve to be studied.

The rest of this paper is organized as follows: Section 2 introduces the main tools used for this research study, namely, the phase-space detection algorithm, the *K*-means clustering technique and the characteristic indices for analyzing the P-wave morphology, the section concludes with the definition of a protocol for P-wave morphology analysis that combines the explained tools. Section 3 starts with a brief study about the performance phase-space algorithm, then the protocol for extracting and classifying the P-waves morphology is applied to 89 ECG signals and the results of this study are presented and discussed later in Section 4. Finally, some conclusions are summarized in Section 5.

2. Description of materials and methods

2.1. Data acquisition

The surface ECG was sampled during normal heart rhythm (sinus rhythm) using modified Frank leads (*X*, *Y* and *Z*) at a rate of 1 kHz and a resolution of 0–625 μV, using a special software and a data acquisition board (equipment supplied by Siemens-Elema AB, Solna, Sweden). The *X*-lead was applied to the fourth intercostal space in both mid-axillary lines; the *Y*-lead at the sternal manubrium, just below the clavicle, and to the left of the fifth of the umbilicus; and the *Z*-lead at the fifth intercostal space and on the spinal backbone.

The ECG data collection was obtained at the Cardiology Department of the Lund University. The number of studied ECGs was 89, most of them being recorded from healthy people. The patients were at rest during the record of 6 min.

2.2. Phase-space detection algorithm

The time evolution of a deterministic autonomous dynamic system can be described in discrete time by a difference equation,

$$x(t+1) = F(x(t)), \quad x(t) := x(t \cdot \tau_s) \in \mathbb{R}^n. \tag{1}$$

where τ_s denotes the sampling time. The state vector $x(t)$ of dimension n, contains the system historical information in discrete time ($t \in \mathbb{Z}$). Usually, the observation of a dynamic system is a discrete signal $s(t)$ that can be defined as a function of the state variables of the system $s(t) = \phi(x(t))$.

In the general case, the dynamics of an n-dimensional dynamic system is not accessible to an observer that usually records a scalar time series of finite size $\{s(t)\}_{t=1}^{L}$. The common method to reconstruct the real dynamics is to embed the time series $\{s(t)\}$ into an N-dimensional vector space [18]. The delay coordinate embedding method [14,2] is the most important phase-space reconstruction technique from a sequence of observations. The vectors of the new embedding phase-space are formed from time delayed values of the scalar measurement signal $s(t)$,

$$X(t) = \begin{bmatrix} s(t-(N-1)\delta) \\ s(t-(N-2)\delta) \\ \vdots \\ s(t) \end{bmatrix}, \tag{2}$$

where N is the embedding dimension and δ is the delay.

The characteristic points of an ECG wave, e.g. P-wave, are related to the wave morphology. The embedding space applied to EGG signals transforms the ECG points to vectors that incorporate additional information about other points in their neighbourhood, see [22]. These vectors can be used to find the characteristic points with an appropriated election of the delay and the embedding dimension. The idea is to search for similar points to a reference in the embedding phase-space.

The ECG signals are usually corrupted by high frequency noise and low frequency disturbances producing baseline oscillations. The high frequency noise can be attenuated by choosing an optimal delay that filters it. On the other hand, the low frequency oscillations can also be reduced by using incremental phase-space variables, i.e., $\Delta s(t) = s(t) - s(t-\delta)$ instead of $s(t)$. In the incremental space, the baseline oscillations with a period of several cycles are removed. Points that have a similar relative position on ECG waves of the same class (with the same morphology) are close in the embedding phase-space, when time is disregarded. The minimization of the distance between points in the embedding phase-space is the basis of the algorithm used in this paper.

The phase-space detection algorithm obtains characteristic ECG points, e.g. the onset and the end of the P-wave, using the proximity of these points in the incremental embedding phase-space. Each point of the ECG signal $s(t)$ is transformed into a vector in the embedding phase-space $X(t)$ with information concerning not only the point but also a sequence of points in its neighbourhood. Finally, the algorithm computes the points in the embedding space that are closer to a reference one in the Euclidean norm.

Let $\mathcal{S}$ denote the signal space corresponding to observations $s(t)$ and let $\mathcal{P}$ denote the transformed phase-space using the embedding approach. The phase-state vector $X(t) \in \mathcal{P}$ associated

to a signal measurement $s(t) \in \mathcal{S}$ is obtained as follows

$$X(t) = \begin{bmatrix} s(t+\beta_1\delta) - s(t+\beta_2\delta) \\ \vdots \\ s(t+\beta_N\delta) - s(t+\beta_{N+1}\delta) \end{bmatrix} \quad (3)$$

where $\beta = [\beta_1, \ldots, \beta_{N+1}]^T$ is a vector of $N+1$ integer numbers that depends on the selected reference point. Good choices for β are the following: $\beta_o = [0, \ldots, N]^T$ when the measurement point is the onset of a wave, $\beta_e = [-N, \ldots, 0]^T$ for the end of a wave, and $\beta_p = [-N/2, \ldots, N/2]^T$ for a peak. Thus, the phase-space vector $X(t)$ associated to the signal measurement $s(t)$ and obtained with β_o includes information about the N next measurements of the signal $s(t)$. The choice of N and δ is a function of the selected wave width. A rule of thumb for the choice of the embedding dimension and the delay values is as follows: the product $N \times \delta$ must approach the wave width in order to include the relevant information about the whole wave morphology in $X(t)$. An illustration of this rule is depicted in Fig. 1. Here the P-wave width is approximately 180 ms, the number of points to encode the wave is selected as $N = 6$, then the value $\delta = 30$ ms is obtained in such a way that $N \times \delta = 180$ ms. Since the point of interest is the P-wave onset, the reference point is the first one in the wave and $\beta = [0, \ldots, 6]^T$. If the point of interest would be the P-wave end, then the same N and δ parameter values could be used, but the reference point would be the last one in the P-wave and β would be selected as $\beta = [-6, \ldots, 0]$.

The Euclidean distance between two points in the phase-space $\mathcal{P}$ represents a morphology distance in the signal space $\mathcal{S}$. The morphology distance from a point $s(t)$ (indexed by t) to a given reference point $s(t_{\text{ref}})$ (indexed by t_{ref}) is denoted as $d(t, t_{\text{ref}})$ and is defined as follows:

$$d(t, t_{\text{ref}}) = ||X(t) - X(t_{\text{ref}})||_2. \quad (4)$$

In order to develop the phase-space detection algorithm and easily visualize its results, the inverse of the distance is a more convenient metric. In such a case, points $s(t)$ that are close to the reference $s(t_{\text{ref}})$ in the morphology distance produce peaks in the plot of the inverse distance. In addition, the inverse distance is modified to avoid division by zero when $t = t_{\text{ref}}$ as follows

$$d^{\dagger}(t, t_{\text{ref}}) = \begin{cases} (d(t, t_{\text{ref}}))^{-1}, & \text{if } t \neq t_{\text{ref}}, \\ M & \text{if } t = t_{\text{ref}}, \end{cases} \quad (5)$$

where M is a large number, such that $M \geq \max_{t \neq t_{\text{ref}}} d^{\dagger}(t, t_{\text{ref}})$.

The peaks (maxima) of $d^{\dagger}(t, t_{\text{ref}})$ indicate the nearest points (in morphology) to the reference one t_{ref}. The shape of the signal $d^{\dagger}(t, t_{\text{ref}})$ is a good test for analyzing the algorithm's accuracy. If the values of δ, N and β have been appropriately chosen, then the function $d^{\dagger}(t, t_{\text{ref}})$ must be multimodal, but with only one clear peak per beat. This means that only one point per beat is adjacent to the reference point.

The phase-space algorithm has been implemented in a very flexible way. Thus, the user can select the reference point on any wave of the recorded ECG in a first stage, and then the algorithm detects all the points that are morphologically close for the rest of the heart beats. This is very convenient for an analyst that has to delineate different parts of the ECG.

2.3. Cluster analysis

Several techniques can be used to classify data into a finite number of classes with similar properties, see [4,10]. Clustering methods classify data by using a data distance. The data are represented as a collection of vectors $\mathcal{S} = \{x_1, \ldots, x_L\}$ in a finite dimension feature space $\mathcal{S} \subset \mathbb{R}^N$ and the data distance can be any vector norm in that space, usually the Euclidean norm of $\mathbb{R}^N$. The K-means clustering method is one of the simplest procedures for solving the problem of classifying L data points $\{x_1, \ldots, x_L\}$ in K different clusters C_k, $k = 1, \ldots, K$. Each cluster is characterized by a centroid that is the barycenter of all the points in the cluster, i.e. the point that minimizes the sum of the distances of all the points of the cluster to it. The method aimed at minimizing an objective function, in this case the squared error function

$$J = \sum_{k=1}^{K} \sum_{j \in C_k} ||x_j - \mu_k||^2 \quad (6)$$

where C_k is the k th cluster and $||x_j - \mu_k||^2$ is the square of the distance from the data point $x_j \in C_k$ to the centroid μ_k of the cluster C_k. A summary of the K-means clustering algorithm is given in Table 1. It has been proved that the algorithm converges, however the computed solution need not be the global optimum. In fact, the algorithm is quite sensitive to the initial locations of the centroids and could be run multiple times with different initial centroids in order to reduce this effect.

The K-means clustering technique has been used as a tool for classifying the different ECG signals in classes depending on the P-wave morphology. In order to accomplish the morphology classification, each P-wave is coded as a vector $x_i \in \mathcal{S}$ using the complete signal from the onset to the end point.

2.4. P-wave morphology indices

The P-wave morphology analysis can be used to predict several atrial pathologies, such as atrial fibrillation, see [19,20,7]. These authors use the Frank leads X, Y, Z and the spatial magnitude SM, defined as $\text{SM} = \sqrt{X^2 + Y^2 + Z^2}$, to predict pathologies. In addition,

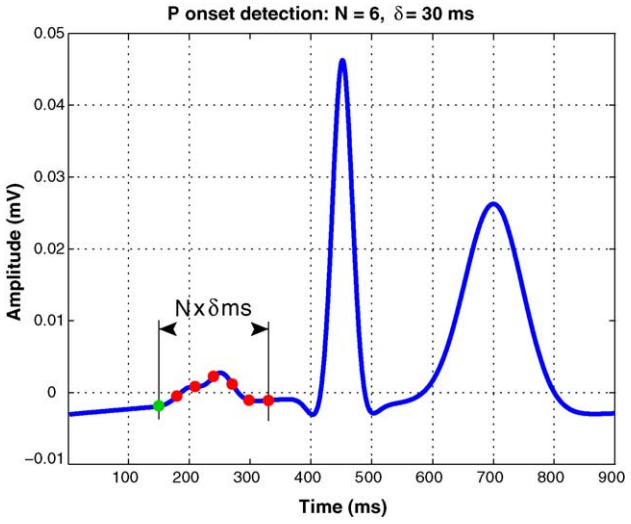

Fig. 1. P-wave onset detection for $N = 6$ and $\delta = 30$ ms. The green point is the reference point and the red ones the additional points to reconstruct the phase-space vector $X(t)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 1
K-Means clustering algorithm.

1	Place the initial group of centroids $\{\mu_1, \mu_2, \ldots, \mu_K\}$ in the space $\mathbb{R}^N$.
2	Assign each object x_i, $i = 1, \ldots, L$ to the group that has the closest centroid.
3	Recalculate de position of the centroids by computing the barycenter of each cluster $\mu_i = \min \sum_{j \in C_i} ||x_j - \mu_i||^2$, $i = 1, \ldots, K$.
4	Repeat steps 2 and 3 until the centroids no longer change.

they define some indices in order to show that the delay in the interatrial conduction of the P-wave is a predictor of some atrial pathologies. The most commonly used indices to analyze the P-wave morphology are: its width, the time of the maximum X and Y peaks, the Z cross-zero time, the minimum and maximum Z peak magnitudes and the peaks of the spatial magnitude SM. For example, a large maximum Z peak and the presence of two peaks in the SM signal may indicate an interatrial delay.

A new index that measures the degree of variability of the ECG P-wave morphology is proposed here. The index of change is defined as follows:

$$\mathrm{IC}(\%) = 100 \cdot \frac{1}{N_B - 1} \sum_{i=1}^{N_B - 1} |\Delta p(i+1)|$$

where N_B is the total number of heart beats in the registered ECG, $p(i) = k$ where k is the index of the P-wave morphology cluster C_k that classifies the present heart-beat, and $\Delta p(i) = p(i) - p(i-1)$.

2.5. A protocol for extracting and classifying the P-wave morphology

The phase-space algorithm and the K-means clustering approach have been combined to define a protocol to extract and classify the morphology of the P-waves contained in a medium to long duration ECG. The developed protocol is shown in Table 2. In the two first steps of the protocol, the onset and end of the P-waves are obtained by using the phase-space detection algorithm, see the details in [12]. In the first step, the Q points belonging to a sinus QRS complex are found. The associated embedding phase-space was obtained by using a delay $\delta = 30$ ms and $\beta = [0, \ldots, 5]$. In the second step, the embedding phase-space for P-wave onset and end detection is obtained with a delay $\delta = 20$ ms, and $\beta_{Po} = [0, \ldots, 5]$ and $\beta_{Pe} = [-5, \ldots, 0]$, respectively. The sinus Q points obtained in the first step are now used as an indication to locate the P-wave onsets and ends. These points are located at the maxima of the inverse distance function and before the previously computed sinus Q points.

The values of the delay δ and embedding dimension N were selected in such a way that the delay δ is sufficiently large to attenuate the high frequency noise and the product $\delta \times N$ is large enough to include the complete wave width of the corresponding characteristic point to be detected. The algorithm is quite robust against small changes in these parameters, as is shown in [12]. Fig. 3 shows a detail of an ECG for a real case (case_3) and the Q, P-wave onset and end points obtained by the detection algorithm.

In the third step of the protocol of Table 2, the P-waves are extracted from the ECG, filtered with a FIR filter, and rotated to obtain normalized P-waves with zero value for the onset and end points.

In the fourth step, the information from the X, Y, Z leads and the SM signal is joined into a vector that comprises the whole P-wave duration. Then, the K-means clustering algorithm classifies these vectors into two groups. The objective of this binary classification is deciding if there exists more than one P-wave morphology. Each cluster represents an average P-wave morphology. If both clusters are close enough, which can be computed using the Euclidean norm, then there is a unique true P-wave morphology.

Table 2
Protocol to detect and classify the P-wave morphologies.

1	Find the sinus Q points using the embedding phase-state algorithm.
2	Find the P-wave onset and end points associated to sinus Q points using the embedding phase-space algorithm.
3	Extract, filter and rotate the P-waves.
4	Classify the P-waves into two clusters using the K-means.
5	Obtain the cluster means and the dynamics between clusters.

Finally, the information about the morphology occurring at each heart-beat is stored in a vector of the same dimension than the number of recorded heart beats. This vector takes on two values, 1 or 2, representing if the corresponding P-wave belongs to cluster C_1 or C_2, respectively.

3. Results

3.1. Performance study of the phase-state detection algorithm

A thorough study about the performance of the phase-space detection algorithm for simulated and real ECGs has been accomplished in [12]. Due to space limitations, only four cases are presented here. The objective is the detection of the P-wave onset of a real ECG for different values of the algorithm parameters. The results are depicted in Fig. 2. The algorithm parameters for the first case are $\delta = 20$ ms, $N = 8$ and $\beta_o = [0, \ldots, N]$. The parameters δ and N are chosen in such a way that their product $\delta \times N = 160$ ms approaches the width of the P-wave and $\beta = \beta_o$ because an onset point is searched. In the second case, the parameters are changed to $\delta = 40$ ms and $N = 4$, using again $\beta = \beta_o$ and the same reference point. The error (given as mean and standard deviation for all the detected points) with respect the points obtained in the first case is -0.022 ± 0.204 ms, that is less than the sample-time of the ECG signal, 1 ms. In the third case, the same reference point and algorithm parameters as in the first case are used, however instead of the three-leads signal only the signal of one lead is used in the detection. The error (mean and standard deviation) with respect to the points obtained in the first case is 0.0074 ± 0.165 ms, again less that the sample-time. Finally, in the fourth case the same algorithm parameters as in the first case are used for the three-lead signal, but the reference point is changed. The error (mean and standard deviation) is now 12.14 ± 0.7555 ms. This means that the new reference point is delayed 12 ms with respect to the first case but the standard deviation of the error is less than the sample-time too. Note that the algorithm does not use time information for detecting the points, but the distance to the reference point in the embedded state-space. In every case, the inverse distance function presents a unique clear peak per cycle that can be easily extracted using either a global threshold or a two-step localized algorithm as previously described. This example shows that the algorithm is very accurate and robust whenever its parameters are chosen in such a way that $N \times \delta$ is approximately the wave width and β is adequately chosen depending on the reference point to be detected.

3.2. A study on the P-wave morphology

A study of the P-wave morphology has been carried out on 89 cases, including patient and control subjects. The protocol of Table 2 has been applied to extract and classify the P-wave morphology information.

This study shows that 18 out of the 89 cases have significant changes in the P-wave morphology. Moreover, the subjects suffer of abrupt changes, but slow in frequency, between the two different P-wave morphologies.

Figs. 3–5 show the results of the application of the extraction and morphology classification protocol for case_3 of our study. Fig. 3 shows a detail of the ECG for case_3 and the Q and P-wave onset and end points obtained by the detection algorithm. In the third step of the protocol of Table 2, the P-waves are extracted from the ECG, filtered by a FIR filter, and rotated to obtain normalized P-waves taking on zero value in the onset and end points. Fig. 4 shows a detail of the result of this process, where the X, Y, Z signals and the spatial magnitude SM are shown. In this figure, all the information of the ECG has been removed,

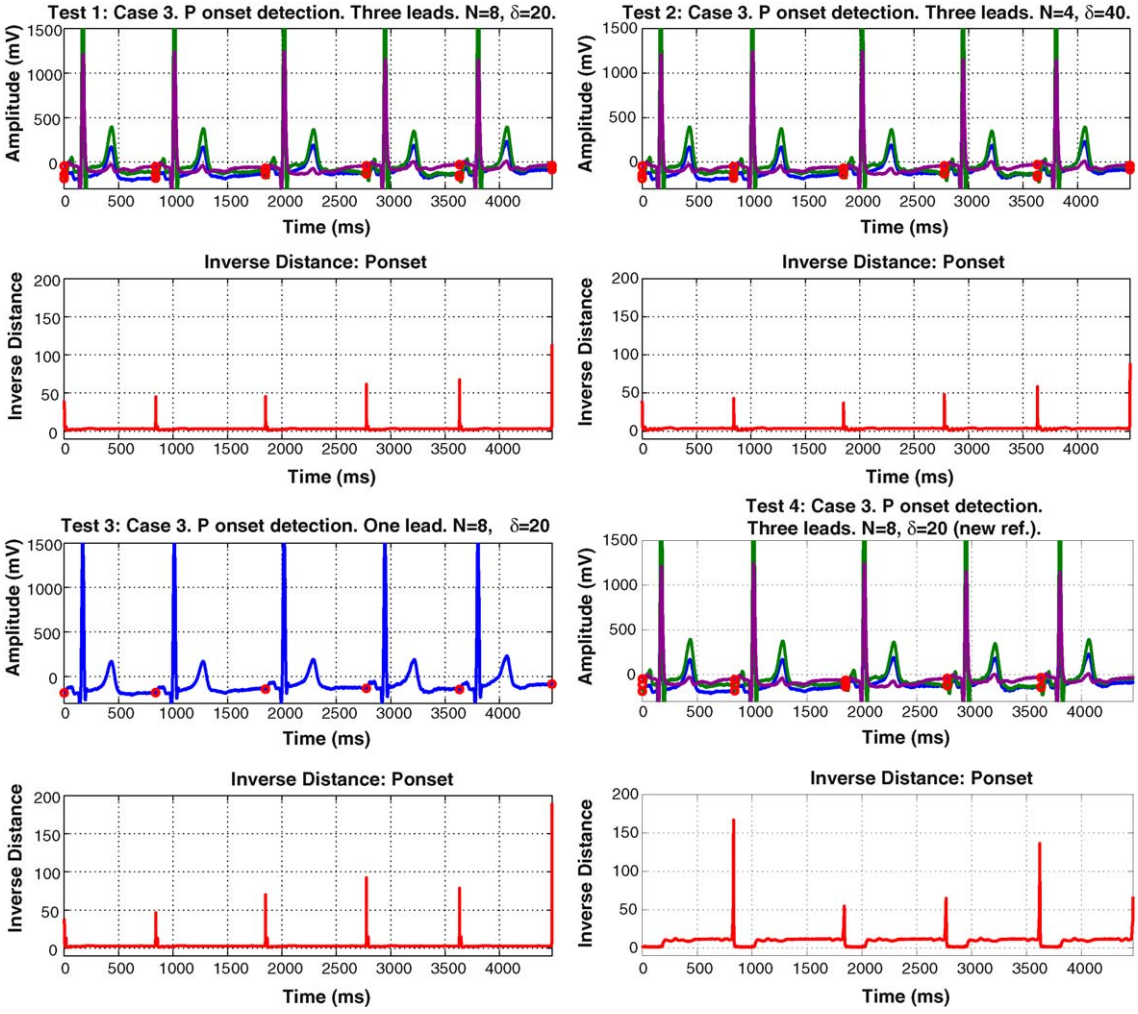

Fig. 2. Performance study of the phase-space detection algorithm. Detail of the ECG signal and inverse distance for four different cases.

except the normalized P-waves for each beat-heart, that are beat-to-beat orderly represented, one after the other. Fig. 5 shows all the P-waves and how they are classified into two distinct clusters.

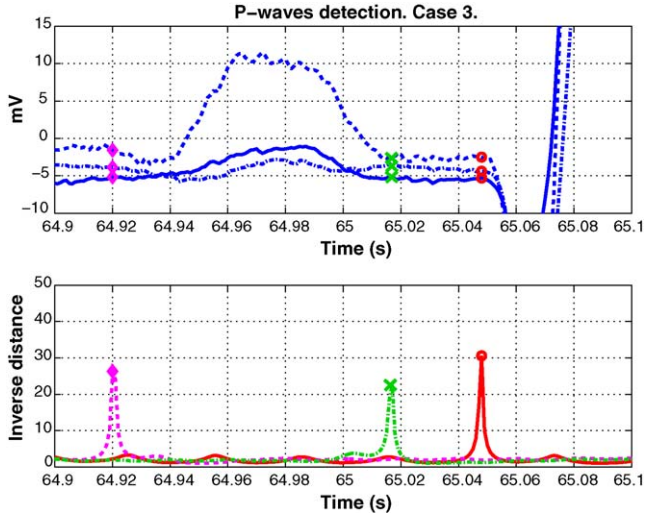

Fig. 3. Detection of P-wave onset, P-wave end and Q event for case$_3$. The upper figure represents the ECG in the three leads. The lower figure represents the modified inverse distance functions for the detection of each event. P-wave onset is marked by a rhombus, P-wave end by a cross and Q event by a circle.

The results of this morphology study for three different cases are shown in Figs. 6–8 . In these figures, the P-wave average clusters after the binary classification process are shown in the normalized *X*, *Y*, *Z* and SM signals. In addition the dynamics of the

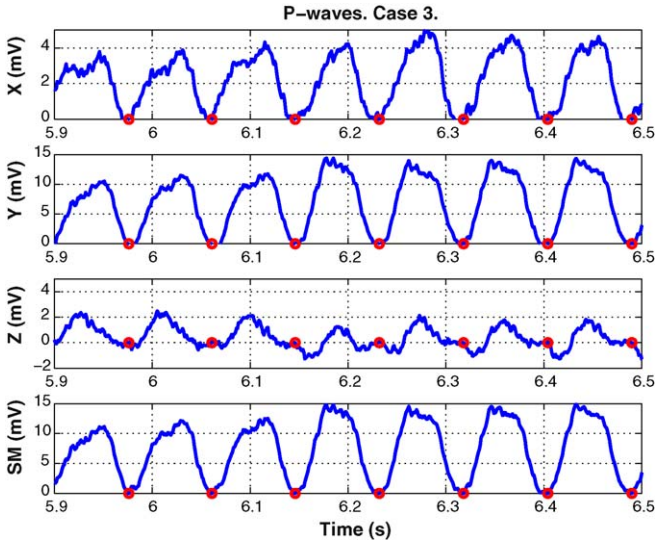

Fig. 4. P-waves detected, isolated, filtered and represented one after the other for case$_3$.

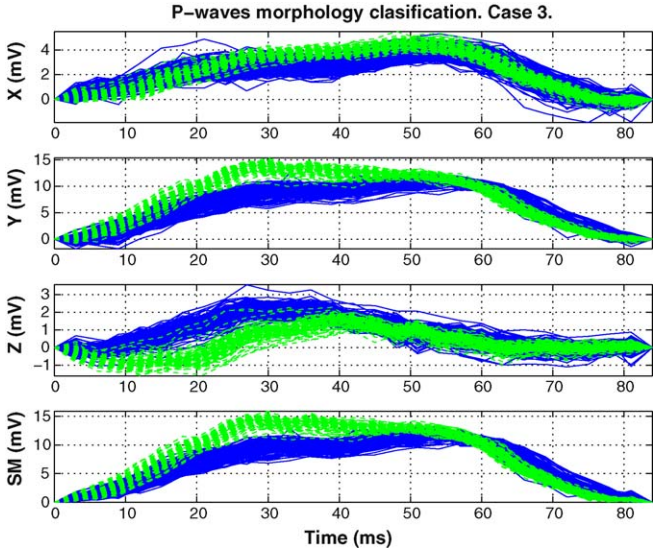

Fig. 5. Superposition of every P-wave of the complete ECG and classification into two clusters for case$_3$. P-waves belonging to the first cluster are depicted in solid line, P-waves belonging to the second cluster are depicted in dashed line.

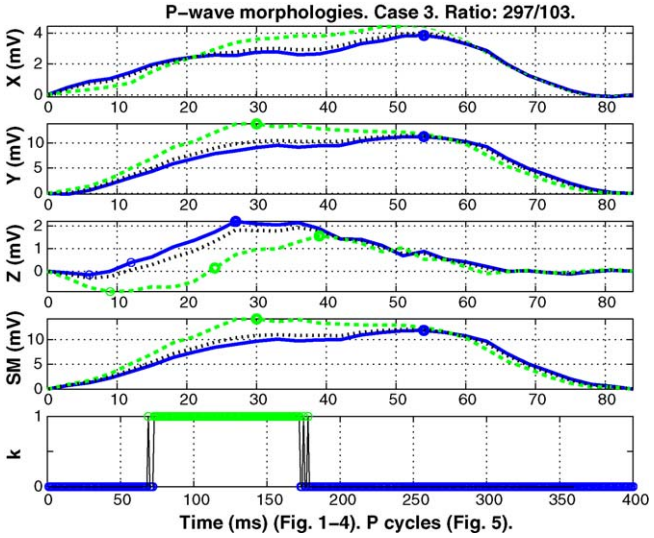

Fig. 7. P-wave cluster means and their dynamic evolution for case$_3$. Different morphologies, slow beat-to-beat changes.

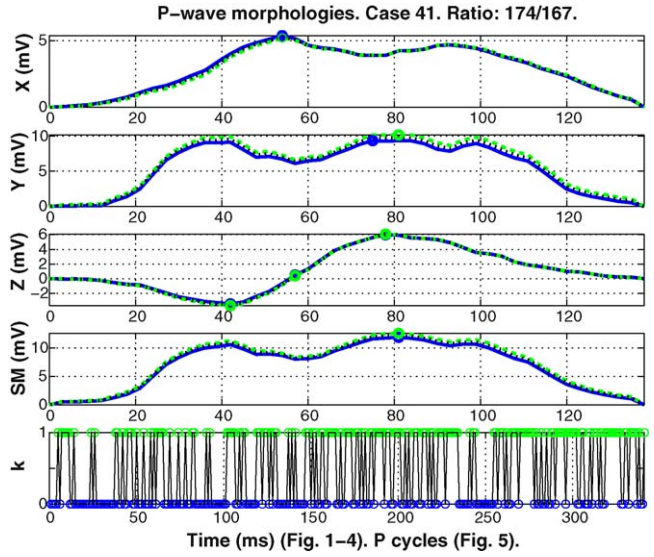

Fig. 6. P-wave cluster means and their dynamic evolution for case$_{41}$. Similar morphologies, fast beat-to-beat changes.

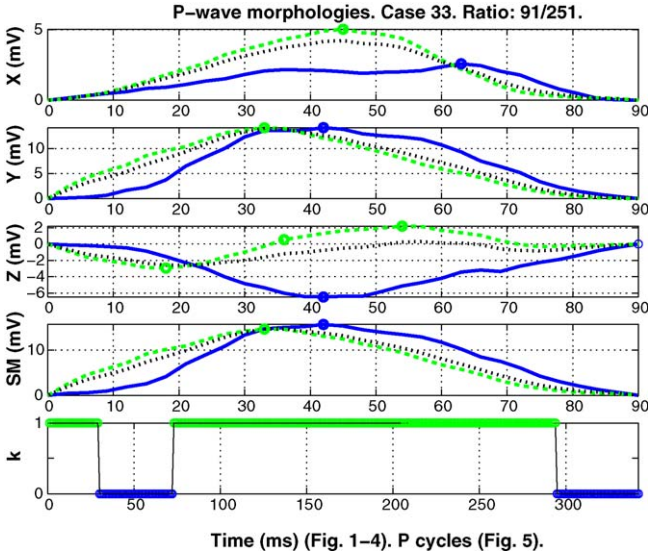

Fig. 8. P-wave cluster means and their dynamic evolution for case$_{33}$. Different morphologies, slow beat-to-beat changes.

beat-to-beat morphology change is depicted in the lowest plot. In this plot, the presence of cluster C_k is represented by its index value k. Note that the heart-beat is represented as an integer number in the axis of abscissas. Two almost indistinguishable P-wave clusters are found for case$_{41}$, as can be checked in Fig. 6. This is the usual case for most of the 89 analyzed cases. Note that fast and almost random changes occur between both almost indistinguishable clusters. This always happens when there is only one true P-wave morphology and corresponds to a high value of the IC index that measures the frequency of change between the obtained P-wave clusters. However, case$_3$and case$_{33}$are essentially different as is shown in Figs. 7 and 8. Two clearly different P-wave morphologies appear in these cases. In addition, the dynamics of the morphology change is very slow, and each P-wave morphology prevails during a large number of heart beats. For instance, the first 75 heart beats in case$_3$corresponds to the first P-wave morphology marked with value 1, the next 100 beats correspond to the second morphology, marked with value 2, then the P-wave morphology changes again to the first one until the end of the registered heart beats. The IC index takes on a small value in these cases. Consequently, there are two different ways to determine if there exists more than one true P-wave morphology. If either the Euclidean distance between the two obtained clusters is small or the IC index is high then there exists only one true P-wave morphology, otherwise there exist at least two different P-wave morphologies. In our study, it has been checked that IC $< 30\%$ for all the cases with two clearly distinct morphologies, and IC $> 80\%$ for all the cases that only have one morphology. Thus, the value IC $= 50\%$ can be used as discriminant for the presence of one or two distinct P-wave morphologies.

Table 3 shows the 18 cases containing two different P-wave morphologies. The tabulated information for each case comprises: the PR width, the P-wave width, the cluster ratio, the X peak and Z cross-zero time positions, the Z and SM maximum magnitudes and the index of change IC. The PR and P-wave widths are expressed in milliseconds (ms), the mean and standard deviation for all the average P-waves is tabulated independently of the morphology

Table 3
P-wave morphology indices for the cases with two morphologies.

Case	PR width (ms), mean ± std	P-wave width (ms), mean ± std	Cluster ratio, C_1/C_2	X peak position (ms), C_1/C_2	Z cross. position (ms), C_1/C_2	Z max (mV), C_1/C_2	SM peak (mV), C_1/C_2	IC (%)
1	124.4 ± 0.8	110.8 ± 0.4	336/37	63/66	60/111	0.4/0.0	16.6/18.0	24.3
2	154.0 ± 2.4	141.7 ± 7.4	48/317	93/66	69/60	4.6/3.6	18.2/9.8	4.2
3	117.8 ± 0.7	84.4 ± 0.5	106/294	51/54	24/12	1.6/2.2	14.2/11.9	6.6
4	192.7 ± 2.7	153.9 ± 2.6	295/62	81/69	66/72	2.0/2.2	19.7/6.8	7.3
7	217.6 ± 5.6	134.7 ± 5.6	74/254	87/66	75/78	4.3/5.3	22.5/17.7	2.7
18	171.0 ± 1.1	123.7 ± 1.1	125/276	75/72	57/51	5.3/6.0	16.1/11.8	4.8
20	151.8 ± 0.8	115.0 ± 0.7	128/185	78/54	108/102	0.1/0.3	15.6/6.8	7.0
33	161.6 ± 0.6	91.1 ± 5.0	91/251	63/45	90/36	0.0/2.2	15.7/14.7	1.6
34	185.7 ± 9.8	131.8 ± 8.6	360/96	90/57	84/75	2.3/2.2	15.3/6.9	10.9
37	214.1 ± 17.4	129.3 ± 15.4	76/340	90/60	78/84	3.9/2.9	12.4/5.8	7.2
43	113.8 ± 7.6	86.9 ± 0.8	114/213	63/42	63/57	0.7/1.2	11.3/9.1	19.7
47	170.5 ± 8.0	138.5 ± 6.8	276/142	69/90	66/69	2.9/3.1	19.1/19.0	6.3
50	224.5 ± 3.8	151.9 ± 4.2	269/56	105/72	75/84	2.9/2.1	15.7/10.0	11.6
51	166.2 ± 0.6	123.1 ± 0.4	200/158	75/51	57/63	4.8/4.2	18.3/10.7	1.3
67	141.3 ± 0.6	112.3 ± 0.7	225/198	87/72	78/87	0.0/0.3	22.3/13.3	1.8
82	138.8 ± 0.6	122.4 ± 0.8	367/62	51/57	51/51	2.3/1.6	11.5/8.3	29.8
85	156.6 ± 6.0	124.5 ± 5.0	172/251	69/51	48/51	0.9/1.6	13.8/10.0	4.9
87	144.2 ± 5.3	132.2 ± 1.3	222/183	81/75	75/87	2.4/1.3	21.4/15.2	2.2

Table 4
Statistical data of the individuals under study.

Number of morphologies	Age (mean ± std)	Cases (male/female)	Cases (control/patient)
1	56 ± 15	36/35	52/19
2	63 ± 12	8/10	17/1

class that they belong to, because these quantities do not have a significant change between both morphology clusters. The remaining indices are averaged for each different morphology cluster and are expressed in the format ($cluster_1/cluster_2$). The indices corresponding to time positions are expressed in milliseconds (ms) and those corresponding to magnitude of the signal in millivolts (mV). Note that in some cases, these indices show a large difference between both clusters, e.g. $case_3$ has two Z cross-zero time values, 24 ms for the first morphology and 12 ms for the second one. The cluster ratio represents the number of P-waves that are classified as belonging to each cluster C_1 and C_2.

Statistical data about the 89 study cases are given in Table 4 although it should be remarked that the objective of our study is not to draw a medical interpretation of these results.

Most of the cases with two different P-wave morphologies correspond to healthy individuals. However, more clinical research needs to be accomplished in order to understand the causes of the change in the morphology. The protocol developed and reported in this paper establishes a procedure to carry out this research study.

4. Discussion

The phase-space detection algorithm was successfully used in the detection of the P-wave characteristic points. The algorithm is very simple because directly uses the ECG signal in the time domain. This provides very accurate detection results and robustness against noise and variability in the shape of the ECG, as it was shown in [12]. The algorithm is semiautomatic, i.e., the user selects a reference point for a certain heart-beat, then the algorithm detects all the similar points to the reference, producing only one per beat. All the 89 cases are ECG signals of patients at rest and have less variation that the simulated signals used in the tests of [12]. Therefore, this reinforces the confidence in the accuracy of the obtained detection results.

The selection of the delay δ and the embedding dimension N was carried out in order to attenuate the high frequency noise and to include the complete P-wave morphology information in the embedding phase-space vectors, such as it was explained in [12].

The P-waves that have been detected, isolated and extracted for the 89 ECGs by using the phase-state detection algorithm, need to be preprocessed before accomplishing a detailed morphology analysis. The P-waves are filtered and rotated in such a way that their onsets and ends attain zero value. Now, a superposition of all the P-waves clearly shows if there exists beat-to-beat morphology changes for some of the studied ECG signals. After this study, an interesting phenomenon has been observed for 18 out of the 89 analyzed cases. For these 18 ECGs, the P-wave presents alternate morphology changes beat-to-beat. This fact implies two interesting questions: First, are the indices associated to the average P-wave valid for these cases? Second, what is the medical interpretation of these morphology changes?

A binary classification of the P-wave morphologies, using the K-means clustering algorithm, was accomplished in order to answer the first question. Most cases exhibit only one morphology, but 18 of them have alternate morphology changes.

The indices explained in Section 2.4 are computed in most studies for an average P-wave assuming that all of them have a similar morphology, see [19,20,7]. However, our study shows that this is not always the case. For example, Fig. 5 corresponding to $case_3$ of our study demonstrates that the P-wave clearly changes its morphology from beat to beat. The use of the average signal could provide spurious information in the corresponding indices. On the other hand, the change in the P-wave morphology is a new phenomenon that has not been completely studied yet and that could report relevant medical information. Consequently, our proposal is to classify the P-waves using a clustering technique, and compute the indices for each cluster in order to study atrial conduction defects. Otherwise the results would provide erroneous conclusions. In addition, a new index IC that measures the beat-to-beat P-wave morphology changes is proposed. The IC index shows that the morphology changes for the 18 cases exhibiting two distinct P-wave morphologies occurs abruptly but with low frequency.

The most rational explanation of the P-wave morphology change is that several control subjects change their interatrial conduction delay during the ECG recording, in agreement with the P-wave morphologies noted by [19]. A change in the sensor position does not explain this phenomenon, because the P-wave morphology changes are detected in every sensor at the same time, see Fig. 4. Clearly, the circumstance that a stable P-wave morphology may abruptly change to another well defined and

stable morphology is a new and hitherto unexplored finding. The biological explanation for this finding remains to be further clarified as also the relation to health or disease.

The protocol of Table 2 permits one to obtain a figure representing all the information about the variability of the P-wave morphology. Then, the morphology analysis can be easily accomplished by simple visual inspection. Figs. 6–8 are three illustrative examples of this methodology. The protocol could be conveniently modified to study the variability of other ECG waves, such us complex QRS, or T-waves.

5. Conclusions

A systematic procedure for studying the variability of the P-wave morphology has been developed. This procedure exploits the versatility of the phase-space detection algorithm along with the *K*-means clustering algorithm. The alternative so far was a tedious process of manual extraction of P-waves. This procedure can be also useful to study the variability of other ECG waves.

A study on 89 ECGs from patient and control subjects, to detect P-wave morphologies that may predict interatrial delay was carried out using this procedure. The P-wave information was obtained by an algorithm of detection based on the embedding phase-space approach with very good results. This algorithm was also developed by the same authors of this paper.

The study shows that some ECGs exhibit abrupt alternate changes in the P-wave morphology. The P-wave morphology indices detect different electrical activity for each morphology. This fact could be used as a medical predictor of certain pathologies. However, more clinical research is needed in order to develop such a prediction method.

Acknowledgments

This work was supported in part by the national research agency of Spain (CICYT) through the project DPI 2006-14367 and the regional government of Castilla y León through the project VA076A07.

References

[1] H.D.I. Abarbanel, The analysis of observed chaotic data in physical system, Review of Modern Physics 65 (4) (1993) 1331–1392.
[2] H.D.I. Abarbanel, Analysis of Observed Chaotic Data Series, Springer-Verlag, New York, 1996.
[3] A. Bayes de Luna, R. Fort de Ribot, E. Trilla, J. Julia, J. Garcia, J. Sadurni, J. Riba, F. Sagues, Electrocardiographic and vectorcardiographic study of interatrial conduction disturbances with left atrial retrograde activation, Journal of Electrocardiology 18 (1985) 1–13.
[4] J.C. Bezdek, S.K. Pal (Eds.), Fuzzy Model for Patient Recognition, IEEE PRESS, New York, 1992.
[5] C.M. Bishop, Neural Networks for Pattern Recognition, 3rd ed., Clarendon Press, Oxford, Great Britain, 1997.
[6] J. Carlson, R. Havmöller, A. Herreros, P. Platonov, R. Johansson, B. Olsson, Can orthogonal lead indicators of propensity to atrial brillation be accurately assessed from the 12-lead ECG? Europace 7 (September (Suppl. 2)) (2005) 39–48.
[7] J. Carlson, R. Johansson, S.B. Olsson, Classification of electrodiographic P-wave morphology, IEEE Transactions on Biomedical Engineering 48 (4) (2001) 401–405.
[8] F. Censi, G. Calcagnini, C. Ricci, R.P. Ricci, M. Santini, A. Grammatico, P. Bartolini, P-wave morphology assessment by a gaussian functions-based model in atrial brillation patients, IEEE Transactions on Biomedical Engineering 54 (April (4)) (2007) 663–672.
[9] J.C. Daubert, D. Pavin, G. Jauvert, P. Mabo, Intra- and interatrial conduction delay: Implications for cardiac pacing, Pacing and Clinical Electrophysiology (PACE) 27 (2004) 507–525.
[10] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., John Wiley & Sons, 2001.
[11] A.I. Hernández, G. Carrault, F. Mora, Improvement of a P-wave detector by a bivariate classification stage, Transactions of the Institute of Measurement and Control 22 (3) (2000) 231–242.
[12] A. Herreros, E. Baeyens, J. Carlson, R. Johansson, B. Olsson, J.R. Peran, An Algorithm for Phase-Space Detection of the P Characteristic Points, in: Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 2004–2007.
[13] F. Holmqvisti, D. Husser, J.M. Tapanainen, J. Carlson, R. Jurkko, Y. Xia, R. Havmöller, O. Kongstad, L. Toivonen, S.B. Olsson, P.G. Platonov, Interatrial conduction can be accurately determined using standard 12-lead electrocardiography: validation of p-wave morphology using electroanatomic mapping in man, Heart Rhythm 5 (3) (2008) 413–418.
[14] H. Kantz, T. Schreiber, Nonlinear time series analysis, Cambridge Nonlinear Science Series, 1997.
[15] B.U. Köhler, C. Henning, R. Orglmeister, The principes of software QRS detection., IEEE Engineering in Medicine and Biology Magazine 21 (2002) 42–57.
[16] P. Laguna, R. Jané, P. Caminal, Automatic detection of wave boundaries in multi-lead ECG signals: Validation with the cse database, Computers in Biology and Medicine 27 (February (1)) (1994) 45–60.
[17] J.P. Martinez, R. Almeida, S. Olmos, A.P. Rocha, P. Laguna, A wavelet-based ECG delineator: evaluation on standard database, IEEE Transactions on Biomedical Engineering 51 (4) (2004) 570–581.
[18] N.H. Packard, J.P. Crutcheld, J.D. Farmer, R.S. Shaw, Geometry from a time series, Physical Review Letters 45 (1980) 712–716.
[19] P.G. Platonov, J. Carlson, M.P. Ingemansson, A. Roijer, A. Hansson, L.V. Chireikin, S.B. Olsson, Detection of inter-atrial conduction defect with unltered signal-averaged P-wave ECG in patients with lone atrial brillation, European Pacing and Clinical Electrophysiology (EU-ROPACE) 2 (2000) 32–41.
[20] P.G. Platonov, S. Yuan, E. Hertevig, O. Kongstad, L.V. Chireikin, S.B. Olsson, Presence of right atrial conduction disturbance in patients with alone atrial brillation, Scandinavian Cardiovascular Journal 35 (4) (2001) 270–279.
[21] F. Portet, P wave detector with pp rhythm tracking: evaluation in different arrhythmia contexts, Physiological Measurement 29 (January (1)) (2008) 141–155.
[22] M. Richter, T. Schreiber, Phase space embedding of electrocardiograms, Physical Review E 58 (5) (1998) 6392–6398.
[23] L. Sörnmo, P. Laguna, Bioelectrical Signal Processing in Cardiac and Neurological Applications, Elsevier Academic Press, 2005.
[24] C. Vazquez, A. Hernández, F. Mora, G. Carrault, G. Passariello, Atrial activity enhancement by wiener filtering using an artificial neural network, IEEE Transactions on Biomedical Engineering 48 (August (8)) (2001) 940–944.