Integrated Shape Averaging of the P-wave Applied to AF Risk Detection

S Boudaoud[1], H Rix[1], JJ Blanc[2], JC Cornily[2], O Meste[1]

[1] University of Nice-Sophia Antipolis, Lab. I3S, France
[2] Brest University Hospital, France

Abstract

The recognition of patients with Atrial Fibrillation (AF) risk, from body surface ECG, remains an actual problem. The aim of the work is to investigate how the study of the P-wave morphology may assess any clustering algorithm in separating patients with AF risk from healthy subjects. The clustering is done by the k-means algorithm and using a new method: the Integrated Shape Averaging (ISA). According to the results, the method was able to separate patients with AF risk from safety ones with 75 % sensitivity and 70 % specificity on 40 recordings. Furthermore, two families of P-wave patterns were characterized among the patients with AF risk and probably corresponding to electrical manifestations of conduction defects in the left or the right atrium.

1.   Introduction

It is well know that the electrocardiogram (ECG) is the signal representation of the electric activity of the heart collected on the body surface. The P-wave is the part of the recorded signal related to the atrium depolarization. Recently, it was observed that patients prone to attacks of Atrial Fibrillation (AF) present conduction defects in the atria [1]. These conduction defects generate shape differences in P- wave morphology on patient recordings [1], [2]. It is interesting to study these shape differences comparing a group of patients who had AF episodes and a group of safety ones. The final objective should be to use the behaviour of shape variations in the prevention of AF risk. J. Carlson et al. [1] have proposed a parametric method based on system identification to perform the shape clustering task. The main drawback of such a parametric method is the lack of direct physiological interpretation. The aim of this communication is to propose a nonparametric method based on ISA to cluster the P-wave shapes which would be better accepted by physicians [3], [4], [5]. This method is able to do the clustering of the P-waves by the k-means algorithm, using ISA method to compute the class centroids, and to characterize two pattern families among the pathologic ones. The two pattern families represent patients probably presenting a conduction delay in the left atrium or in the right one [2].

2.   Materials and methods

The signals we are interested in for shape clustering are assumed to be positive on their support. In practice this hypothesis is not restrictive for we can either use the energies or separately process the positive and negative parts of the signals [4]. We use the k-means algorithm combining a similarity criterion derived from the Distribution Function Method (DFM) [3] and the ISA algorithm [4] to get centroids [5]. The principles of these algorithms are briefly recalled below.

Distribution Function Method (DFM): The principle is based on the idea that the normalized integrals S(t) and V(t) of the positive signals s(t) and v(t) can be viewed as Distribution Functions and that the shape difference between the two signals is completely characterized by function φ defined by the relation [3]:

$$S(t) = V(\varphi(t)) \tag{1}$$

The departure of $\varphi(t)$ from linearity, for example computing the residue from the least mean square line, is a similarity criterion. In fact in our work, two similarity criterions d_1 and d_2 derived from the DFM are used. Starting with M values of t (t_j, j = 1 to M) and the corresponding $\varphi(t_j)$, we computed:

$$d_{1S,V}(i) = (\frac{1}{i}\sum_{j=1}^{j=i}(\varphi(t_j) - \alpha_i t_j - \beta_i)^2 +$$

$$\frac{1}{M-W-i}\sum_{j=i+W}^{j=M}(\varphi(t_j) - \alpha_i t_j - \beta_i)^2)^{\frac{1}{2}}$$

$$d_{1S,V} = \max_i(d_{1S,V}(i)) \tag{2}$$

where $y = \alpha_i t + \beta_i$ is the equation of the LMS line fitting $\varphi(t_j)$ in function of t_j, for $j = i$ to $i+W$. W is the length of the observation window and $i=1$ to M-W. The second similarity criterion is defined as:

$$d_{2S,V} = (\frac{1}{M}\sum_{j=1}^{j=M}(\varphi(t_j) - \alpha t_j - \beta)^2)^{\frac{1}{2}} \qquad (3)$$

where $y = \alpha t + \beta$ is the equation of the LMS line fitting $\varphi(t_j)$ in function of t_j, for $j = 1$ to $M' < M$.

With d_1, we have no a priori knowledge on shape variations and we search for the similarity criterion that maximises (2). With d_2, we utilize the information that shape distortion is due to the delay between the both components associated to the right and left atria generating a shape asymmetry [1], [2], [5].

Integrated Shape Averaging (ISA): Let $x_i(t)$, for $i = 1$ to N, be N positive signals with normalized integrals $X_i(t)$. For every value y, let us compute N abscissa t_i, solving

$$y = X_i(t_i) \qquad \text{for } i = 1 \text{ to } N \qquad (4)$$

The main idea of ISA is to average these abscissas for each ordinate value y, $0 \le y \le 1$ leading to:

$$\bar{t}(y) = \frac{1}{N}\sum_{i=1}^{i=N}t_i \qquad (5)$$

Plotting y in function of $\bar{t}$ provides us with a distribution function which is the normalized integral of our mean shape signal:

$$y = \widetilde{X}(\bar{t}) \qquad (6)$$

It has been shown [4] in the case where all the signals $x_i(t)$ are random scaled and shifted versions of the same signal $s(t)$ following model (5):

$$x_i(t) = k_i s(\alpha_i t - d_i) + m_i(t) \qquad (7)$$

Where $m_i(t)$ is a zero mean Gaussian white noise realisation.

that the mean shape signal $\widetilde{x}(t)$, obtained by derivating $\widetilde{X}(t)$, can be deduced from $s(t)$ through the average affin function, i.e. whom parameters are the average scale factor and the average delay. So, in a class of equal shape signals, this synthesized signal is a good candidate for the

centroid [5]. In addition, ISA is also used to replace the classical signal averaging to reduce noise on repetitive signals. The main advantage of using ISA is to be insensitive to scale and delay fluctuations [4].

3.   Results and discussions

To test our approach, we used a data set of 20 recordings of 1 min duration sampled at 1 kHz (D2 lead) of patients who had AF episodes and 20 recordings of safety patients. Each recording was segmented by physicians and the baseline removed. For the clustering of the P-wave shapes, we made use of $\widetilde{x}(t)$ instead of the classical averaged signal, both to get a mean signal in a series of about 35 beats for each recording and to compute each class center. An example of averaged signal is given on Fig. 1. We can observe the shape difference between the two signals obtained by both averaging techniques.
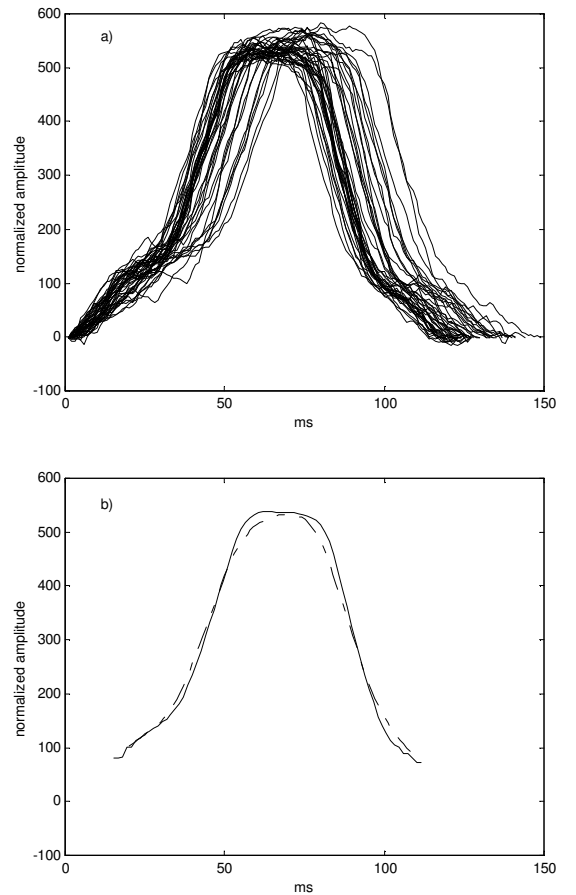

Figure 1. a) Successive segmented P-waves from one recording, and b) the mean signal (continuous line) obtained by ISA and the classical averaged signal (dashed line).

The Clustering Strategy: The clustering operation is done following two clustering strategies. The first one consists of using the similarity criterion d_1 (Eq. 2) and the k-means algorithm assuming 3 classes: safety (S), pathologic 1 (P1) and pathologic 2 (P2) representing patients with conduction defects in the left atrium and the right one respectively. The second strategy is decomposed in two steps. First, using similarity criterion d_2 (Eq. 3), the clustering is performed to separate 2 classes: (S + P2) and P1. Then, the algorithm is used another time with d_2 to separate in two classes the first class of the precedent clustering: S and P2. A time transformation is performed on the signals of this class before this clustering step:

$$p_i(t) = p_i(T_i - t) \quad \text{for i=1 to } N_{(S+P2)} \qquad (8)$$

Where T_i represents the length of the P-wave p_i.

During the clustering, to avoid the influence of an atypical shape inside the class in computing the class center, a similarity test is done with the other class signals. If the result is over a specific threshold, the concerned signal is not used for the class center computation.

In Tables 1, 2 and 3 the obtained results are shown with the two clustering strategies and using the same signals for the class centers initialization. The φ_i functions are computed on 89 points, $W = 45$ and $M' = 36$.

Table 1. Number of classed signals.

Classes	Safety	Path. 1	Path. 2
Strategy 1	20	6	14
Strategy 2	19	11	10

Table 2. Number of well classed signals.

Classes	Safety	Path. 1	Path. 2
Strategy 1	12	6	6
Strategy 2	14	9	6

Table 3. Global clustering performances.

Performances	Sensibility	Specificity
Strategy 1	60 %	60 %
Strategy 2	75 %	70 %

According to the results, we can see that the patients of the class P1 are well detected among all the recording with few errors especially with the strategy 2 of clustering (9 over 11 recordings). The shape differences with safety and P2 recordings are important. On the contrary, the patients of the class P2 are more difficult to detect due to smaller differences in shape with safety patients. For this class, strategy 2 performs better than strategy 1 (6 over 10 recordings). Even with the safety recordings, the results of strategy 2 are better (14 over 19 recordings). We can observe on Fig. 2 the final class centers obtained by strategy 2 for P1, P2 and safety patients respectively.
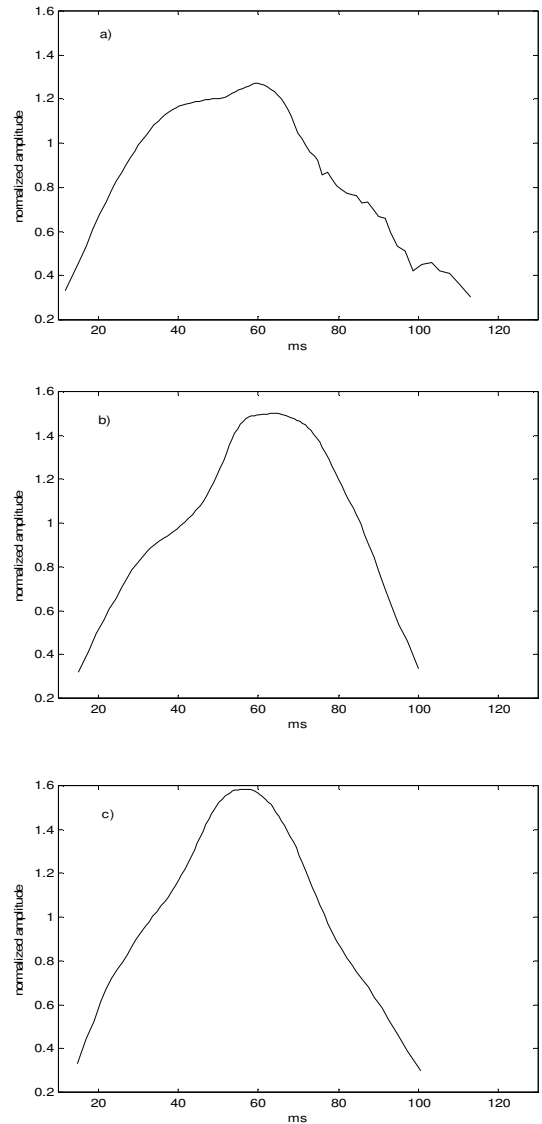

Figure 2. The final class centers for the class P1 (a), P2 (b) and S (c) obtained with the strategy 2 of clustering.

Each pattern represents the mean shape of the concerned class. The P1 pattern is characterized by rapid raise and a slow descent with duration of about 120 ms. On the contrary; the P2 pattern presents a slow rise and a rapid descent with no excessive duration. The safety pattern is characterized by a relative symmetry with rapid rise and descent. All these remarks are relevant with the previous discussions in the literature [2]. The departure from perfect clustering of our data can be explained by: 1) the use of a suboptimal similarity criterion for the clustering task, 2) the presence of more than two pathological patterns inside recordings, 3) a natural shape variability among recordings which makes the shape clustering difficult, and 4) the possibility that some patients presenting conduction defects in the atria had not AF episodes until now.

4.   Conclusion

We have proposed a new shape clustering method based on Integrated Shape Averaging. The method was adapted to characterize P-wave shapes. The clustering results are promising and indicate the possibility of designing similarity criterions for detecting a specific P-wave pattern. This pattern could be linked with a specific AF pathology or an evolution step of it. The clustering goal could be to extract a population with AF risk from a group of patients and to separate this population into distinct pathologic classes. For this purpose, it must be noticed that we showed that P-wave shape is an important feature to detect AF risk but other features may be added to improve the clustering performances.

References

[1] Carlson J, Johansson R, Olsson SB.   Classification of Electrocardiographic P-Wave Morphology. IEEE Trans Biomed Eng 2001;48(4):401-5.
[2] Clavier L, Boucher JM, Blanc JJ. P-wave parameters for atrial fibrillation risk detection. 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 1996;1367-8.
[3] Rix H, Malengé JP. Detecting small variations in shape. IEEE Trans Syst Man Cybern 1980;10:90-6.
[4] Rix H, Meste O, Muhammad W. Averaging signals with random time shift and time scale fluctuations. In Proceedings of BSI 2002;19-22.
[5] Rix H, Boudaoud S, Meste O. Clustering signal shapes: Application to P-Waves in ECG. In Proceedings of EMBEC 2002; 364-5.

Address for correspondence.

Sofiane Boudaoud
Lab. I3S, UNSA-CNRS, 2000, route des Lucioles. BP 121
0603 Sophia Antipolis Cedex , France
tel. / fax: +33 04 92 96 51 04/ +33 04 92 96 51 55

E-mail: Boudaoud@i3s.unice.fr