Prediction of Atrial Fibrillation Termination by Approximate Entropy in the Time-Frequency Domain

C Vayá[1], JJ Rieta[2], R Alcaraz[1], C Sánchez[1], R Cervigón[1]

[1]Innovation in Bioengineering Group, Castilla-La Mancha University, Spain
[2]Bioengineering, Innovation and Development, Universidad Politécnica de Valencia, Spain

Abstract

Given the high prevalence of Atrial Fibrillation (AF) among adult population, to distinguish between these AF episodes that terminate spontaneously and those that persist if no external action is carried out becomes a subject of great clinical interest. In this matter, the complexity analysis of mathematical sequences of parameters obtained from time-frequency distributions of electrocardiograms are useful to recognize the type of AF. The parameters which complexity has been analyzed are the mean peak frequency of atrial activity and the spectral concentration. The complexity measurements of the sequences have been made using the Approximate Entropy, repeating the same procedure for nine distinct time-frequency distributions. The possibility of classifying correctly the AF episodes depends directly on the selected time-frequency distribution.

1. Introduction

Supraventricular tachyarrhythmias, in particular the Atrial Fibrillation (FA), are the most commonly encountered in the daily clinical practice. The prevalence of AF is less than 1% among population under 60 years old, but it increases significantly among those over 70, approximating to 10% in those older than 80 [1]. When the AF terminates spontaneously we refer to it as paroxysmal AF. On the contrary, when some electrical or pharmacologic cardioversion is needed it is called persistent AF [1]. Frequently, persistent AF results in permanent AF [1, 2], which is closely related to a rising probability of suffering embolisms and these might provoke strokes [1]. For this reason, it is important, from a clinical point of view, to distinguish between paroxysmal and persistent AF episodes, because an effective diagnosis based on noninvasive techniques such as electrocardiogram (ECG) could reduce the number of hospitalizations.

Recent research proves that it is possible to use statistical analysis of ECG parameters in the time-frequency domain as a tool to classify AF episodes [3]. An additional step to these studies would be to measure the regularity of time sequences obtained from those parameters and check the applicability to AF classification. In this sense, entropy estimators of numerical sequences such as Approximate Entropy (ApEn) [4, 5] are useful to measure its complexity. These estimators have already been used to characterize some biomedical signals like electroencephalography registrations [6] or fetal ECG studies [7]. The novelty of this paper is to make a entropy study from a previous transformation of the Atrial Activity (AA) signal to time-frequency domain, and the later construction of time sequences of parameters obtained from the time-frequency distributions. The main objective consists of proving the existence of significative regularity differences between sequences from paroxysmal and persistent AF. This estimation is based on two characteristic parameters of the time-frequency distributions: the main peak frequency (f_p) and the Spectral Concentration (SC) of the AA. Calculations have been repeated to nine distinct time-frequency distributions, so that we can test which of them optimize results.

2. Database

We have analyzed a total of 30 ECG signals of one minute in length extracted from 24-hours one-lead Holter recordings of AF patients. The original sampling rate of the Holter systems was 128 samples per second, but ECG signals have been interpolated by a factor of 8, so that the resulting sampling rate f_s equals to 1024 samples per second. In this way we obtain a higher resolution of the ECG signals in the time-domain and in the time-frequency domain, and a better cancelation of the QRS complex. Sixteen of the signals belong to persistent AF patients (N-group), since termination was not observed during the whole observation time of these patients. The rest of the signals are annotated as paroxysmal AF (T-group), given that in all of them the AF episode terminates one second after the end of the one-minute registration.

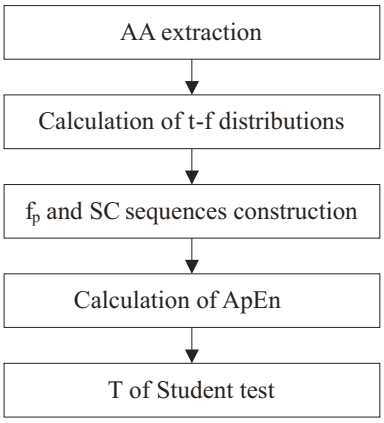

Figure 1. Global process to calculate the complexity of sequences obtained from time-frequency distributions

3. Methods

3.1. Global process

The global process used to obtain the measures consists of five main steps. Firstly, the Atrial Activity (AA) is extracted from the ECG registers, given that the analysis of previously separated AA makes easier the study of AF [8] and improves the information provided by time-frequency distributions [9]. Secondly, we calculate nine distinct time-frequency distributions of the previously obtained AA. Next, the f_p and SC sequences of every time-frequency distribution are constructed. Then the $ApEn$ of these sequences is calculated. Finally, the t of Student test is applied to the obtained $ApEn$ values in order to know whether the level of significance between N and T groups that allows us to distinguish them. The global process is schematized if Fig. 1.

3.2. Extraction of the atrial activity

There exist several techniques designed to extract the AA of AF episodes from ECG registrations. The limitation of having only one-lead ECG obliged us to discard those techniques based on the spatial diversity of multi-lead systems, such as blind source separation [8]. On the contrary, the average beat substraction technique [10] works efficiently with one-lead ECG , so this has been the technique chosen to extract the AA.

3.3. Time-frequency distributions

After extracting the AA from the ECG, we calculate nine time-frequency distributions of the AA. All the calculated distributions belong to the Cohen's class [11], and are the following: spectrogram, Wigner-Ville, pseudo-Wigner-Ville, Margeneau-Hill, pseudo-Margeneau-Hill, Page, pseudo-Page, Zhao-Atlas-Marks and Choi-Williams. All distributions have been calculated with the same resolution of $1sec$ in time and $1Hz$ in frequency. Given that the number of samples in every ECG signal is elevated (approximately equal to 60000), the calculation of distribution has been applied on signal subdivisions of 1024 samples in length. This allows us to reach the aforementioned spectral resolution and, at the same time, the computational load of calculations is considerably reduced.

3.4. Construction of sequences

The next step, after calculating the time-frequency distributions, consist of obtaining from every one of them the time sequences of the f_p and SC parameters. Since a time-frequency distribution can be considered as a sequence of spectra calculated in consecutive moments, to construct the time sequences of f_p and SC we extract their values from the distinct calculated spectra of the time-frequency distribution.

The value of f_p is obtained extracting the frequency at which the main peak occurs in a certain AA spectrum. The second time sequence obtained from the time-frequency distributions is the SC, which tries to estimate the level of energy concentration of the AA around f_p. It is calculated as [12]:

$$SC = \frac{\sum_{f=0.82f_p}^{1.17f_p} P_{AA}(f)}{\sum_{f=0}^{0.5f_s} P_{AA}(f)} \quad (1)$$

where P_{AA} is the power spectrum of the AA signal, f is the frequencies vector, and f_s is the sampling frequency.

3.5. Approximate entropy

The $ApEn$ measures the complexity of the time sequences. It quantifies how predictable a time sequence is depending on how many times similar patterns are repeated [4, 5].

Let $x[n]$ be a time sequence of length equal to N. The distance between two of its subsequences $X_m(i)$, $X_m(j)$ of m length is defined as:

$$d[X_m(i), X_m(j)] = \max(|x(i+k) - x(j+k)|) \quad (2)$$

For a certain $X_m(i)$ we calculate $C_i^m(r)$ as:

$$C_i^m(r) = \frac{n_i^m}{N-m+1} \quad (3)$$

where n_i^m is the number of subsequences that fulfill $d[X_m(i), X_m(j)] < r$ with $1 \leq i \leq N-m+1, j \neq i$,

and r is the parameter that defines the level of likelihood between subsequences [4]. Next $C^m(r)$ is calculated as the mean of these $C_i^m(r)$, that is:

$$C^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} C_i^m(r) \quad (4)$$

Finally, the $ApEn$ is defined as:

$$ApEn(x[n], m, r) = \ln \left\{ \frac{C_m(r)}{C_{m+1}(r)} \right\} \quad (5)$$

4. Results

In table 1 the $ApEn$ results of f_p and SC time sequences are summarized in terms of mean and standard deviation values and organized by time-frequency distribution and patient group (N or T). In addition, these tables also contain the bilateral significance between patients groups that results from applying the t of Student test to $ApEn$. As it can be observed, in three of the time-frequency distributions, the bilateral significance obtained in both f_p and SC parameters is quite low (lower than 0.05), what allows us to reject the hypothesis of equal means and conclude that there exist significative differences between N and T patients groups. Among these three distributions, the spectrogram is the one that reach the least bilateral significance in both f_p and SC cases, remaining lower than 0.001. Also the pseudo-Margeneau-Hill distribution achieves low values of bilateral significance, which is equal to 0.003 and 0.024 for f_p and SC respectively. These same values are obtained by the pseudo-Page distribution. The mean $ApEn$ obtained from the three aforementioned distributions is higher in the N group that in the T group for both f_p and SC parameters, what reveals the greater complexity of the N-group signals. In the rest of the time-frequency distributions, the resulting bilateral significance is too high to consider that a correct discrimination between N and T groups can be done.

In table 2, the areas under the ROC curves of the $ApEn$ of f_p and SC are expressed in percentages for every time-frequency distribution. In the f_p case, the highest area is obtained by the spectrogram (87.9%), and it equals to 80.4% in the pseudo-Margeneau-Hill and pseudo-Page distributions. In the rest of the time-frequency distributions, the area under ROC curves is lower than 60%. With reference to the SC parameter, the highest area under the ROC curve is achieved by the pseudo-Margeneau-Hill and pseudo-Page distributions (83.9%), followed by the spectrogram (72.8%), Wigner-Ville (65.6%) and de la pseudo-Wigner-Ville (61.2). The rest of time-frequency distributions take values lower than 60%.

The Receiver Operating Characteristic (ROC) curves of the distributions that obtained the best results (spectrogram, pseudo-Margeneau-Hill and pseudo-Page) are depicted in Fig. 2 for f_p and SC. A suitable choice of the threshold allow us to distinguish correctly in a high percentage of cases between N and T AF patients. For example, if we pay attention to the f_p ROC curve of the spectrogram we see that it is possible to choose a threshold so that the 81.3% of the N type patients are correctly classified with a false alarm probability equal to 21.4%. With regard to the SC ROC curve of of the pseudo-Margeneau-Hill distribution, we see that it is possible to choose a threshold so that the 87.5% are correctly classified with a false alarm equal to 21.4%.

f_p					
	N group		T group		
	Mean	STD	Mean	STD	Sig.
SP	0.1311	0.0314	0.0788	0.0346	<0.001
WV	0.0470	0.0150	0.0566	0.0264	0.239
PWV	0.0514	0.0182	0.0579	0.0227	0.394
MH	0.0360	0.0054	0.0346	0.0071	0.533
PMH	0.0178	0.0043	0.0121	0.0051	0.003
PG	0.0360	0.0054	0.0346	0.0071	0.533
PPG	0.0178	0.0043	0.0121	0.0051	0.003
ZAM	0.0398	0.0180	0.0375	0.0169	0.727
CW	0.0385	0.0171	0.0352	0.0156	0.576
SC					
	N group		T group		
	Mean	STD	Mean	STD	Sig.
SP	0.4822	0.0743	0.4009	0.0976	<0.001
WV	0.2029	0.0354	0.2297	0.0568	0.142
PWV	0.2358	0.0370	0.2500	0.0501	0.390
MH	0.2256	0.0831	0.2289	0.0675	0.904
PMH	0.1932	0.0445	0.1362	0.0766	0.024
PG	0.2256	0.0831	0.2289	0.0675	0.904
PPG	0.1932	0.0445	0.1362	0.0767	0.024
ZAM	0.1493	0.04718	0.1365	0.050	0.484
CW	0.1443	0.0313	0.1332	0.0349	0.368

Table 1. Mean, standard deviation of f_p and SC, and bilateral significance between T and N groups of spectrogram(SP), Wigner-Ville (WV), pseudo-Wigner-Ville (PWV), Margeneau-Hill (MH), pseudo-Margeneau-Hill (PMH), Page (PG), pseudo-Page (PPG), Zhao-Atlas-Marks (ZAM) and Choi-Williams (CW) distributions

5. Conclusions

The study carried out lead us to conclude that it is suitable to apply complexity measures to sequences obtained from time-frequency distributions in order to classify the AF type as terminating or non-terminating. Three of the studied time-frequency distributions (spectrogram, pseudo-Margeneau-Hill and pseudo-Page) are useful to discriminate between patients groups. The mean $ApEn$ of f_p and SC obtained from these three distributions is higher in the N group that in the T group, thus a greater complexity of the non-terminating AA signals is revealed.

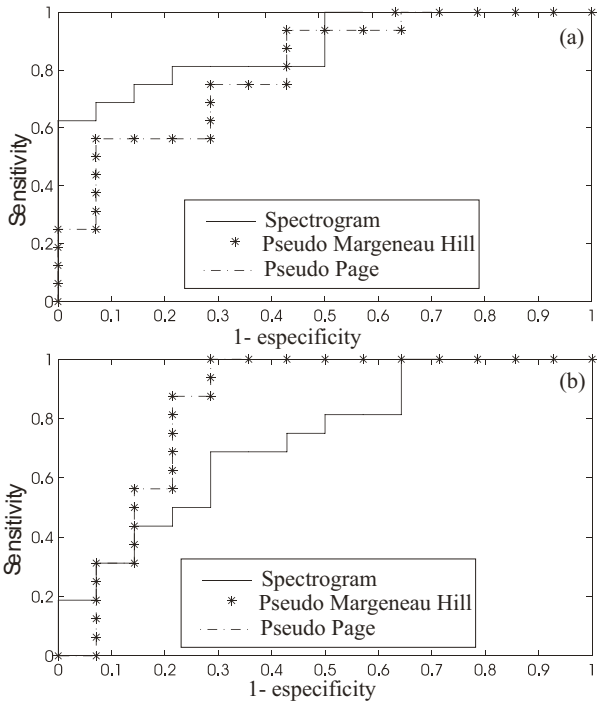

Figure 2. ROC curve of the $ApEn$ of (a) f_p and (b) SC obtained from spectrogram, pseudo-Margeneau-Hill and pseudo-Page distributions

	fp	SC
ESP	87.9%	72.8%
WV	58.0%	65.6%
PWV	56.3%	61.2%
MH	57.8%	50.1%
PMH	80.4%	83.9%
PG	57.8%	50.2%
PPG	80.4%	83.9%
ZAM	50.4%	58.0%
CW	53.6%	58.0%

Table 2. Area under the ROC curves of the $ApEn$ of f_p and SC sequences obtained from spectrogram (SP), Wigner-Ville (WV), pseudo-Wigner-Ville (PWV), Margeneau-Hill (MH), pseudo-Margeneau-Hill (PMH), Page (PG), pseudo- Page (PPG), Zhao-Atlas-Marks (ZAM) and Choi-Williams (CW) distributions.

The complexity study of other time-frequency distribution parameters and the application of other complexity estimators are possible subjects of future research.

Acknowledgements

This work was partly funded by the research project PAC-05-008-1 from *Consejería de Educación de la Junta de Comunidades de Castilla-La Mancha* and the project GV06/299 from *Consejería de Empresa, Universidad y Ciencia de la Generalitat Valenciana*.

References

[1] Fuster V, Ryden LE, Asinger et al RW. CC/AHA/ESC guidelines for the management of patients with atrial fibrillation. Journal of the American College of Cardiology 2001;38(4):1266/I–1266/LXX.

[2] Wijffels M, Kirchhof C, Dorland R, Allessie M. Atrial fibrillation begets atrial fibrillation. A study in awake chronically instrumented goats. Circulation 1995;(92):1954–1968.

[3] Mora C, Castells F, Ruiz R, Rieta JJ, Millet J, Snchez C. Prediction of spontaneous termination of atrial fibrillation using time frequency analysis of the atrial fibrillatory wave. In Proc.Int.Conf on Computers in Cardiology. 2004; 109–112.

[4] Pincus SM. Approximate entropy (apen) as complexity measure. Chaos 1995;5(1):110–117.

[5] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;(278):H2039–H2049.

[6] Hornero R, Abasolo D, Jimeno N, I SC, Poza J, Aboy M. Variability, regularity and complexity of time series generated by schizophrenic patients and control subjects. IEEE Trans Biomed Eng 2006;53(2):210–218.

[7] Ferrario M, Signorini G, Magenes G, Cerutti S. Comparison of entropy-based regularity estimators: Application to the fetal heart rate signal for the identification of fetal distress. IEEE Trans Biomed Eng 2006;53(1):119–125.

[8] Rieta JJ, Castells F, Sánchez C, Zarzoso V. Atrial activity extraction for atrial fibrillation analysis using blind source separation. IEEE Transations of Biomedical Engineering 2004;51(7):1176–1186.

[9] Holm M, Pehrson S, Ingemansson M, Srmno L, Jahansson R, Sandhall L, Sunemark M, Smideberg B, Olsson C, Olsson SB. Non-invasive assessment of the atrial cycle length during atrial fibrillation in man : Introducing, validating and illustrating a new ECG method. Cardiovascular Research 1998;38:69–81.

[10] Slocum J, Sahankian A, Swiryn S. Diagnosis of atrial fibrillation from surface electrocardiogram based on computer-detected atrial activity. Journal of Electrocardiology 1992; 25(1):1–8.

[11] Auger F, Flandrin P, Gonalvs P, Lemoine O. Time-Frequency Toolbox for use with MATLAB. Centre National de la Recherche Scientifique, 1996.

[12] Castells F, Rieta JJ, Millet J, Zarzoso V. Spatiotemporal blind source separation approach to atrial activity estimation in atrial tachyarrhythmias. IEEE Transactions on Biomedical Engineering 2005;52(2):258–267.

Address for correspondence:

Carlos Vayá
Innovation in Bioengineering.
Universidad de Castilla-La Mancha.
Camino del Pozuelo, s/n, 16071, Cuenca.
carlos.vaya@uclm.es