Journal of Arrhythmia 32 (2016) 57–61

Contents lists available at ScienceDirect

Journal of Arrhythmia

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

Original Article

Development of an alert system for subjects with paroxysmal atrial fibrillation

R.A. Thuraisingham, Ph.D. (Cambridge, UK)[1,*]

1A, Russell Street, Eastwood, NSW 2122, Australia

ARTICLE INFO

Article history:
Received 29 July 2015
Received in revised form
17 August 2015
Accepted 24 August 2015
Available online 3 November 2015

Keywords:
Paroxysmal atrial fibrillation
RR interval
Spectral entropy
Standard deviation

ABSTRACT

Background: Knowledge of the onset of atrial fibrillation (AF) episodes in patients with paroxysmal atrial fibrillation (PAF) will enable them to better manage this condition. Current advances in mobile technology allow RR interval data to be obtained in real time. An analysis technique using RR interval data is presented with a view to alert a subject before a PAF episode.

Method: The method is based on a time series of standard deviation and 0.99 quantile values of the spectral entropy, constructed from RR data. The RR data are taken from three time periods. The first time period has no occurrences of AF for 45 min to either side of the time period. The second time period just precedes an AF attack. Both of these are of thirty minutes duration. The third time period of approximately 5 min follows the second, and is when AF occurs.

Results: Twenty-two PAF subjects were studied and in all cases there was a steady increase in the values of these indices as the onset of the AF attack approached.

Conclusion: This method of analysis of RR interval data shows potential use to alert a PAF subject before the onset of an AF episode.

1. Introduction

Atrial fibrillation (AF) is the most common major cardiac arrhythmia and affects a large number of people, particularly the elderly [1]. Paroxysms of AF sometimes precede the onset of more sustained AF. Early detection of paroxysmal atrial fibrillation (PAF) in electrocardiography will enable the patient to obtain timely treatment to manage and prevent further complications [2]. In this communication, a method using RR intervals is presented which signals the onset of PAF.

Heart rate data, unlike ECG traces, are easily monitored using many commercially available wearable devices. They are used mainly by those engaged in strenuous physical activity. It is the purpose of this communication to determine the feasibility of using heart rate data to examine whether they could be used to alert a PAF subject before a sudden attack of AF. Such a warning system would help a PAF subject to seek preventive treatment immediately before such an attack occurs.

Before we proceed to develop such a system, let us examine the results of some relevant studies which involved analysis of RR intervals from PAF subjects. The RR interval is the inverse of the heart rate. In a study of a large group of PAF patients [3], RR interval data collected in the hour preceding PAF and the twenty minutes before PAF were divided into four 5-min period, and were analyzed. It was observed that there was a significant increase in the standard deviation (SD) of the RR intervals from 65 ± 4 to 70 ± 4 ms ($p < 0.02$) before the onset of PAF. The symbol p is the probability of observing the result under the assumption of the null hypothesis. There was also a linear decrease in mean RR interval from 925 ± 16 to 906 ± 16 ms ($p < 0.0002$) before the onset of PAF. In the frequency domain, there was a significant increase in high-frequency (HF) components before PAF ($p < 0.001$), and a decrease in low-frequency (LF) components ($p < 0.0001$). The low/high frequency ratio also showed a linear increase until 10 min before PAF, followed by a sharp decrease immediately before PAF. Another study on the onset of arrhythmia using Holter monitoring [4] also showed higher values in the SD of RR intervals in the last 5-min period before the start of PAF, compared with the other intervals. The analysis was conducted using six 5-min periods before the onset, and a control 5-min period. In the frequency domain there was significant lowering in the LF region in the last 5 min before onset, while the HF was

* Tel.: +61 2 98743022.
E-mail address: ranjit@optusnet.com.au

[1] An Honorary Research fellow with the Rehabilitation Studies Unit Northern Clinical School, University of Sydney, NSW 2065, Australia.

http://dx.doi.org/10.1016/j.joa.2015.08.006
1880-4276/

similar in all intervals, as well as in the control period. Dimmer et al. [5], in their analysis of PAF, also observed an increase in the SD of RR intervals in the last 5 min before AF occurred. However, heart rate variability analysis in the frequency domain did not reveal any significant changes before the onset of AF.

On the other hand, RR interval analysis by Vickman et al. [6] in 22 PAF subjects, which involved using segments of 20 min, showed no significant changes in the traditional time and frequency variability indices. However, their study showed that nonlinear HR variability measures, the approximate entropy, and the altered fractal properties (short term scaling exponent α), showed significant changes before the onset of PAF. There was a decrease in the approximate entropy, from 1.09 ± 0.26, 120 to 100 min before AF, to 0.88 ± 0.24, 20 to 0 min before AF, with $p < 0.001$; in the short term scaling exponent α, there was a decrease from 1.01 ± 0.28, 120 to 100 min before AF, to 0.89 ± 0.28, 20 to 0 min before AF. It is possible that their inability to observe the changes in SD, HF, and LF could be due to the choice of large segment length used in this study.

Examination of the results of all of the above studies showed that, except for the work of Vickman et al. [6], the SD of the RR interval showed a significant increase when an AF attack was imminent. Furthermore, the observation of Vickman et al. [6] that nonlinear changes accompany the onset of AF is a hint to search for a nonlinear index, which is suitable for short segments of data. This led to the analysis of RR data in this paper to be carried out using SD of the RR intervals, a linear time domain index, and spectral entropy (SE) [7], a non-linear index.

SD and SE are measures of variability and disorder of a time series, respectively. SE is a novel measure of the disorder of a time series. Approximate entropy (ApEn) [8] and Sample entropy (SampEn) [9] are some of the indices that have been used to measure the disorder of a system. However, unlike ApEn or SampEn, the evaluation of SE does not require a large amount of RR interval data, which is a distinct advantage in the development of a fast response warning system. SE is a measure of the disorder in the frequency domain of a time series, and is evaluated using power spectra. Although the work here on SE uses some of the parameter choices from the work of Staniczenko et al. [7], not all are the same. The use of the randomness property of the RR intervals in this study was prompted by the fact that the distribution of the RR intervals exhibit random behavior in the AF region [10]. Since SE is not commonly used in RR interval analysis, a brief summary of its evaluation is given in Section 2. In this paper, both SD and SE are examined with the view that they will complement each other and provide confidence to the PAF subject that an AF episode is imminent.

Let us now briefly describe the manner in which the data are analyzed in this paper. The data used in this study were obtained from the public internet site Physionet [11]. RR data from 22 PAF subjects are used in this study [12]. The RR interval data were collected from three different time periods. The RR interval data in the first time period of thirty minutes, referred to as A in this paper, represent the time period during which there are no instances of PAF, for 45 min on either side of that time range. The RR intervals in the second time period of thirty minutes, referred to as B in this paper, represent the time period just preceding an AF attack. Immediately following time period B, an episode of AF occurs. This third period, referred to as C in this paper, when PAF occurs, has a duration of about five minutes. In this communication, the RR interval data in time periods A, B, and C are concatenated before analysis. This is done with a view to determine whether one could observe changes in SD and SE in sufficient time so that the PAF subject can take appropriate action before the onset of PAF. Both SD and SE are examined in this paper instead of alone, with the view that they will complement each other and provide confidence to the PAF subject that an AF episode is imminent.

The paper is structured as follows. Section 2 gives a brief description of the SE method, the data used, and the analysis procedure. Section 3 describes the results, and Section 4 is the conclusion.

2. Method and materials

2.1. Evaluation of spectral entropy

The evaluation of SE of the RR interval time series is carried by first symbolizing the RR data. One problem that is encountered in obtaining the power spectra of the RR data is that the time intervals between beats are not uniform. This is overcome by using a constant sampling interval and symbolizing the RR data, where the RR data are converted into a binary string [7]. Initially a suitable sampling interval (τ), a window length (L samples), and a window separation length (Ls samples) are chosen. Each of the sampling intervals present in a window is categorized as either 1 or 0, depending on whether it contains a beat or not. Thus, each window contains a uniformly-sampled time series of 1's and 0's. After obtaining the binary string, the power spectrum of this segment of data is obtained using the discrete Fourier transform (DFT) with a rectangular window.

If the power of the ith frequency is C_i then the probability of having that power can be defined as

$$p_i = \frac{C_i}{\sum_{i=1}^{L/2} C_i} \tag{1}$$

where the number of frequency bins is $(L/2)$, assuming L is even and the mean is removed.

The entropy of this probability distribution referred to as the SE is then given by

$$H = \sum_{i=1}^{L/2} -p_i \log_2 p_i \tag{2}$$

White noise has power at all frequencies, and has a flat power spectrum. Such a spectrum gives the maximum value for the SE:

$$H_{max} = \log_2(L/2) \tag{3}$$

In order to have the range of SE in $[0, 1]$, the value of H in Eq. (2) is normalized by H_{max} (Eq. (3)).

The choice used for the sampling interval (τ), the window length (L), the shape of the window function in DFT, and window separation length (Ls) was guided by the choices of Staniczenko et al. [7]. The sampling interval was chosen by the criterion that there are no useful frequencies above $(1/2\tau)$. Such a criterion led to the choice of the sampling interval of 0.03 s [7,13]. The choice of the window length L differed from that chosen by Staniczenko et al. [7]. They sought the shortest window length to give sufficient resolution in the frequency domain. This was achieved with the window length being chosen to contain at least 10 beats. In the present study, window lengths with different beats were examined. It was found that a window length with an average between 40 and 60 beats provided the best distinction between different regions of the RR data. Finally, a window length of 40 beats was chosen. Since the number of beats in the window is small compared to the length of samples in the segment, the shape of the window was chosen to be rectangular. This window shape has a low value for the equivalent noise band width (ENBW) [14,15]. In order to obtain a large number of SE values, windows were overlapped. The shape and the overlap between windows are the same

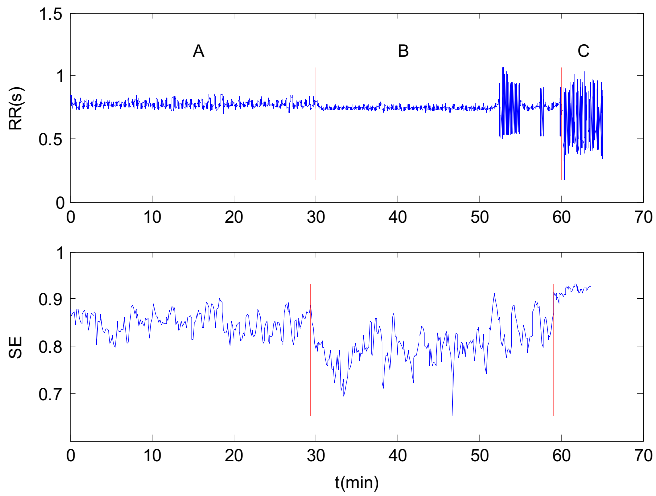

Fig. 1. (Top) Plot of RR intervals versus their time of occurrence; (bottom) plot of SE as a function of time. The red lines demarcate the regions A, B, and C.

as that chosen by Staniczenko et al. [7,14,15]. The overlap between windows was set at 0.75L, giving a window separation $L_s = L/4$.

2.2. Data used

The study employs the learning data set made available in the Cardiology 2001 Challenge [11] and found in the Physionet data bank [12]. The records used for this study are those of patients with PAF. The sex, age, and medical history of the subjects, and whether data were obtained during the day or at night, were not made available to the public. For each subject, two 30-min samples and one 5-min sample are used. The first 30-min sample is distant from any PAF episode, while the second 30-min sample directly precedes an episode. The third 5-min sample is a continuation of the second record, when the PAF episode occurs. The first sample is distant from any PAF episode, and has no PAF episodes within 45 min on either side of the excerpt. The RR intervals were extracted using the software from the Physionet toolbox [16]. The RR intervals obtained from such a heart rate time series were further de-trended and ectopic beats removed [17] before being used for analysis. Ectopic beats were removed since they introduce bias in heart rate analysis, and the origins of some are not completely physiological [18]. Fig. 1 shows the RR interval data for all three periods. The vertical lines separate the periods A, B, and C. The focus of this paper is to develop a procedure that will alert the PAF subject before the onset of the attack in C.

2.3. Analysis procedure

In order to study and develop a warning system to alert a PAF subject before the onset of an AF attack, the RR interval time series of regions A, B, and C were concatenated, so that each subject was associated with just one RR interval time series. Region A is further away from any AF attack, region B precedes the attack, and region C is when the AF attack occurs. In the manner described in Section 2.1, a time series of SE samples are evaluated for each RR interval time series. Suppose SE (i) and t (i) are the SE and the time in minutes corresponding to each window. Let us digress at this point, and consider the evaluation of the SD time series, which is much easier, before a procedure is outlined how these SE values can be used in the development of an alert system.

Let us consider an initial time interval containing nr_M RR intervals, starting from the first RR interval in region A, and let sd (1) be the SD in the time interval that contains the range of RR intervals from [0, nr_m]. An SD time series is then constructed {sd (1), sd(2),…, sd(nl)}, where sd(k) corresponds to the SD in the time interval that contains the range of RR intervals [0 ($nr_M + k - 1$)]. Thus a time series of SD values is constructed where the first SD corresponds to nr_M RR intervals, followed by a series of SD values corresponding to ($nr_M + 1$), ($nr_M + 2$)…, RR intervals. Thus, the difference in the SD values between successive terms in the series arises from a single RR interval. Let nr_A, nr_{A+B} and nr_{A+B+C} be the number of RR intervals in regions A, (A+B), and (A+B+C) respectively. The last term sd(nl) in the SD series corresponds to the SD containing nr_{A+B+C} RR intervals. Now that a time series of SD values is evaluated, let us consider a similar type of series, using the sequence of SE values that were evaluated before.

Let SE(ns_M), SE(ns_A), SE(ns_{A+B}), SE(ns_{A+B+C}) be the SE sample that is closest to the time corresponding to nr_M, nr_A, nr_{A+B} and nr_{A+B+C} RR intervals. Using the set {SE(1), SE(2)…, SE(ns_M)}, various statistical indices were evaluated, namely the mean (SE_{mn}(1)), 90% quantile value (SE_{90}(1)), and 99% quantile value(SE_{99}(1)). Subsequent samples $SE_{mn}(i)$, $SE_{90}(i)$, $SE_{99}(i)$ were evaluated using the set {SE(1), SE(2),…, SE($ns_M + i - 1$)}, $i > 1$. Thus, a time series similar to SD is obtained for SE_{mn}, SE_{90}, and SE_{99}. In the latter three series, the separation in time is the same as SE.

In each of the time series of SD, SE_{mn}, SE_{90}, and SE_{99}, the sample size of the RR intervals, and SE values that are used to evaluate the samples in these series, increase with time. Thus, unless there is a marked increase in the RR and SE values with time, one would not expect the values in the constructed time series to show an increasing trend. These time series, which reflect the time variation of variability and disorder of the RR interval series, will be tested using the data described in Section 2.2, to determine their usefulness in providing prior warning to a PAF subject of an impending AF attack.

3. Results

The number of RR intervals nr_M to construct the time series was chosen to ensure that the sample size was sufficient to obtain a reliable estimate of sd(1), SE_{mn}(1), SE_{90}(1), SE_{99}(1), and that the number of samples for these series in region A are not small. Numerical experiments showed that a choice of 300 RR intervals in region A was sufficient. This corresponds to an approximate time interval of 4 min and nearly 15% of the RR intervals and SE samples found in region A. After constructing the time series SD, SE_{mn}, SE_{90}, SE_{99}, the difference in the mean values of these samples in regions A, (A+B), and (A+B+C) were evaluated for all 22 subjects. This was done to determine whether the mean value in region (A+B) is greater than in region A, and the mean value in region (A+B+C) is greater than in region (A+B). If this is so, then one would expect that there is a steady increase in these indices with time. Such a trend would make these indices suitable markers to indicate an impending AF attack. After evaluating the mean values of these indices in the three regions for the 22 subjects, a paired t-test was carried to determine whether the differences were statistically significant. The p-value showed that SD and SE_{99} provided the best statistic, where the mean value in region (A+B) was greater than in region (A), and the mean value in region (A+B+C) was greater than the mean value in region (A+B+C). The p-value for SE_{mn} showed that the differences between region (A+B) and A were statistically insignificant (>0.05), while for SE_{90} it was >0.01. For differences between region (A+B+C) and (A+B), the p-values for SE_{mn} and SE_{90} were much better; the p-value for SE_{mn} was >0.01 and for SE_{90} it was 2.8(-4). However, in comparison, for SE_{99} and SD, the p-value for the differences between region (A+B) and A were 0.0011 and 3.7(-5), respectively; for differences between (A+B+C) and (A+B), the p-values for SE_{99} and SD were 1.3(-6) and 1.5(-7), respectively. A

Table 1
Summary of paired *t*-test results.

Regions compared	Statistic used	*p*-Value	% Success
(A+B) and A	SE_{99}	0.0011	86.4
(A+B+C) and (A+B)	SE_{99}	1.3(−6)	100
(A+B) and A	SD	3.7(−5)	100
(A+B+C) and (A+B)	SD	1.5(−7)	100
(A+B) and A	$SE_{99}(a)$	2.9(−5)	100
(A+B+C) and (A+B)	$SE_{99}(a)$	3.6(−5)	100
(A+B) and A	$SD(a)$	3.2(−5)	100
(A+B+C) and (A+B)	$SD(a)$	2.3(−7)	100

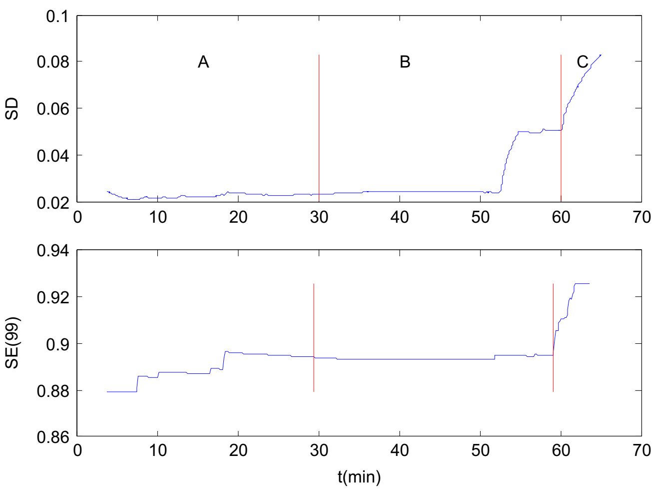

Fig. 2. (Top) SD versus time; (bottom) SE_{99} versus time. The red lines demarcate the regions A, B, and C. The plots shown here use the data of Fig. 1.

summary of these results for SE_{99} and *SD* is given in Table 1. For SD, the values in region (A+B) were greater than region A, and values in region (A+B+C) were greater than in region (A+B) in all 100% of the PAF subjects; for SE_{99}, this occurred in 86% and 100% of the PAF subjects, respectively. These results prompted this study to focus only on the two indices, SD and SE_{99}, in the development of an alert system for PAF subjects.

In Fig. 2, a time series plot of SD and SE_{99} is shown using the data shown in Fig. 1. The results indicate an increase in the values of the variables as a function of time. In order to overcome the discontinuities that occur, so as to obtain a smoother variation and thus make it easier to observe the underlying increase in the values of these variables, an approximation of these variables was obtained using wavelet analysis [19]. The Daubechies wavelet 'db4' is used and the approximation is constructed at level 10. Such an analysis can be carried out as a function of time once the time series is available. Fig. 3 shows the results of this analysis for the time series SD and SE_{99} given in Fig. 2. The approximations are referred to as $SD(a)$ and $SE_{99}(a)$. Wavelet analysis allows the approximation to be plotted simultaneously with that of SD and SE_{99}.

Using these wavelet approximations, the difference in the mean values of these samples in regions A, (A+B), and (A+B+C) were reevaluated for all 22 subjects. The results showed that for SD, the *p*-value for the mean value in region (A+B) to be greater than in region (A) is 3.2(−5), and the *p*-value for the mean value in region (A+B+C) to be greater than the mean value in region (A+B) is 2.3(−7). For SE_{99}, the corresponding *p*-values are 2.9(−5) and 3.6(−4). Furthermore, for these two approximations of SD and SE_{99}, all 100% of the subjects showed values in region (A+B) to be greater than in region A, and values in region (A+B+C) to be greater than in region (A+B).The results of the statistical tests for the wavelet approximations are also summarized in Table 1.

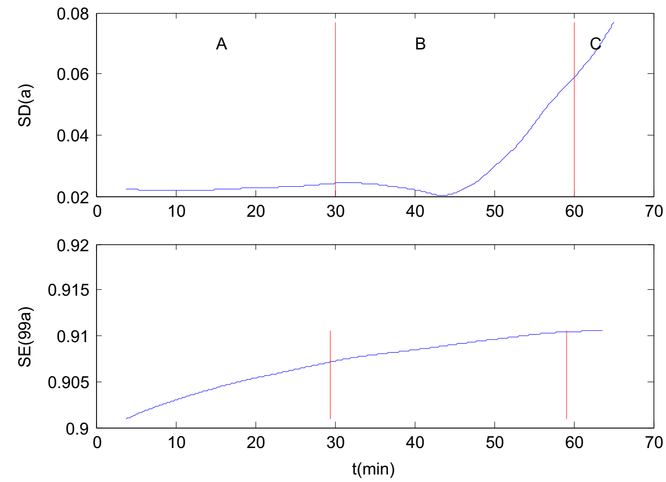

Fig. 3. Plot of the approximation from wavelet analysis as a function of time. (Top) $SD(a)$ and (bottom) $SE_{99}(a)$. The red lines demarcate the regions A, B, and C.

The results show that when an AF attack is imminent for a PAF subject, both SD and SE_{99} begin to increase. The wavelet approximations for these time series provide a smoother variation. Thus, both SD and SE_{99} are useful indices to alert a PAF subject of an impending AF attack, and to seek appropriate treatment. The increase in these indices just before an AF attack is seen clearly in all the PAF subjects studied here. This is encouraging, and can provide a basis to develop an alert system for PAF subjects.

4. Discussion

This study was carried out in order to develop an early warning system to alert a PAF subject before the onset of a sudden AF attack. With the availability of mobile technology to obtain heartbeat data in real time, such a method would benefit PAF patients greatly. This study used RR interval data sets collected in three time periods from normal sinus rhythm in PAF subjects. One data set is far removed from any PAF episode, while the other directly precedes it. The third set is a continuation of the second set, and is the period when a PAF episode occurs. The method employed for this analysis is based on the use of SD, a linear measure of variability, and SE, a non-linear index, which measures the disorder of the time series. In the latter case, the SE is not used directly, but instead a time series constructed from the 0.99 quantile values of the SE is used. So that the method is easily adaptable for a continuous real-time application, the time series of these indices are constructed not using constant distinct intervals, but where the indices are evaluated over intervals that change continuously with time. Both these properties are evaluated separately as a function of time, using the concatenated RR interval series in the three periods. In order to obtain a smooth variation of these values as a function of time, wavelet analysis was carried out to obtain an approximation free of details. The result is an increasing value in these properties as the time of a PAF attack approaches. The study shows that although either of the properties can be used for detection, the use of both will provide more confidence to the PAF subject as to when to seek help. The evaluation of SE and the quantile values are computationally fast, and provide additional evidence of the PAF attack. The successful prediction in 100% of the subjects studied here is encouraging.

The implementation of this methodology for practical use would involve the rapid transfer of the heartbeat data into a suitable device, where the RR data are preprocessed and analyzed. Such a device can be either a part of the RR collection unit or separate. Since the method assumes that there is an initial period of about 4 min where there is no PAF, there is a time lag of about 4 min before the results of the analysis are seen. This time lag of 4 min appears to be the only limitation, since after this the SD and SE_{99} values are evaluated after each sample of RR and SE. This makes it a suitable procedure in real time.

Although the technique developed here to signal the onset of PAF shows much promise, the study was limited by the data set used. Since the data used were from the public domain, personal details of the subjects, in particular their medical history and data collection history, were not available. Furthermore, the sample set is not large. This makes the study less exhaustive. However, the positive results observed here should be an impetus for further studies to be undertaken with data sets having more samples and known subject and data collection history.

5. Conclusion

This study shows that a time series constructed using the SD and the 0.99 quantile values of the SE using heartbeat data offers a feasible method to alert a PAF subject of an imminent AF attack. The method was tested using RR interval data from 22 PAF subjects, and the results indicate that in all cases there was a steady increase in these indices as the onset of the AF attack approached. This signature of a steady increase in these values provides a useful marker to alert the PAF subject to seek help. With advances in mobile technology that allow RR interval measurements to be made in real time from heartbeat data, this study demonstrates that the use of this technique in the analysis of RR data could provide much-needed help for PAF patients to better manage the onset of episodes.

Conflict of interest

All authors declare no conflict of interest related to this study.

References

[1] Feinberg WM, Blackshear JL, Laupacis A, et al. Prevalence, age distribution and gender of patients with atrial fibrillation. Arch Intern Med 1995;155:469–73.
[2] Prakash A, Saksena S, Hill M, et al. Acute effects of dual-site right atrial pacing in patients with spontaneous and inducible atrial flutter and fibrillation. J Am Coll Cardiol 1997;29:1007–14.
[3] Zimmermann M, Bettoni M. Autonomic tone variations before the onset of paroxysmal atrial fibrillation. Circulation 2002;105:2753–9.
[4] Vincenti A, Brambilla R, Fumagalli MG, et al. Onset mechanism of paroxysmal atrial fibrillation detected by ambulatory Holter monitoring. Europace 2006;8:204–10.
[5] Dimmer C, Szili-Torok T, Tavernier R, et al. Initiating mechanisms of paroxysmal atrial fibrillation. Europace 2003;5:1–9.
[6] Vickman S, Mäkikallio TH, Yli-Mäyry S, et al. Altered complexity and correlation properties of R–R interval dynamics before the spoe aneous onset of paroxysmal atrial fibrillation. Circulation 1999;100:2079–84.
[7] Staniczenko PPA, Lee CF, Jones NF. Rapidly detecting disorder in rhythmic biological signals: a spectral entropy measure to identify cardiac arrhythmias. Phys Rev E 2009;79:011915.
[8] Pincus SM. Approximate entropy as a measure of system complexity. Proc Natl Acad Sci USA 1991;88:2297–301.
[9] Richman JS, Moorman JR. Physiological time series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;278: H2039–49.
[10] Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Phys Rev Lett 2002;89:0681021–4.
[11] Goldberger AL, Amaral LAN, Glass L, et al. Physiobank, physiotoolkit and physionet: components of a new research resource for complex physiologic signals. Circulation 2000;101:e215–20.
[12] Circ Challenge. Data sets. The PAF prediction challenge data base [online]; 2001 〈http://www.physionet.org/physiobank/database/afdb〉.
[13] Ng J, Goldberger JJ. Understanding and interpreting dominant frequency analysis of AF cardiograms. J Cardiovasc Electrophysiol 2007;18:680–5.
[14] Harris FJ. On the use of windows for harmonic analysis with the discrete Fourier transform. Proc IEEE 1978;66:51–83.
[15] Nuttall AH. Some windows with very good side lobe behavior. IEEE Trans Acoust Speech Signal Process 1981;29:84–91.
[16] 〈http://www.physionet.org/physiotools/〉.
[17] Thuraisingham RA. Preprocessing RR interval time series for heart rate variability analysis and estimates of standard deviation. Comput Methods Progr Biomed 2006;83:78–82.
[18] Peltola MA. Role of editing RR intervals in the analysis of Heart rate variability. Front Physiol 2012;3:148.
[19] Misiti M, Misiti Y, Oppenheim G, et al. Wavelet Toolbox for use with Mathlab. Math Works Inc 1996;1-31:2–17.