Available online at www.sciencedirect.com
ScienceDirect
Medical Engineering & Physics 31 (2009) 116–123
Medical Engineering & Physics
www.elsevier.com/locate/medengphy

A probability density function method for detecting atrial fibrillation using *R–R* intervals

Lu Hong-wei *, Sun Ying, Lin Min, Li Pi-ding, Zheng Zheng

College of Medical Instrumentation and Food, University of Shanghai for Science and Technology, China

Received 30 May 2007; received in revised form 29 April 2008; accepted 30 April 2008

Abstract

A probability density function (PDF) method is proposed for investigating the structure of the reconstructed attractor of *R–R* intervals. By constructing the PDF of distance between two points in the reconstructed phase space of *R–R* intervals of normal sinus rhythm (NSR) and atrial fibrillation (AF), it is found that the distributions of PDF of NSR and AF *R–R* intervals have significant differences. By taking advantage of their differences, a characteristic parameter k_n, which represents the sum of n points slope in filtered PDF curve, is put forward to detect both 400 segments of NSR and AF *R–R* intervals from the MIT-BIH Atrial Fibrillation database. Parameters such as number of *R–R* intervals, number of embedding dimensions and slope are optimized for the best detection performance. Results demonstrate that the new algorithm has a fast response speed with *R–R* intervals as short as 40, and shows a sensitivity of 0.978, and a specificity of 0.990 in the best detecting performance.

Keywords: Atrial fibrillation (AF); *R–R* intervals; Probability density function (PDF); Phase space reconstruction

1. Introduction

Atrial fibrillation (AF) is a cardiac arrhythmia that causes the heart to beat irregularly, resulting in inefficient pumping of the blood and ultimately blood clots and strokes. AF is responsible for 15–20% of all strokes and is predicted to be associated with over 3 million hospitalizations by 2025 [1]. Since AF can be asymptomatic or paroxysmal, ambulatory monitoring is an ideal choice for detecting it. However, the electrocardiogram (ECG) collected by Holter or other ambulatory monitors may contain significant muscle artefacts [1] or other noise, which could possibly blur the weak characteristic wave of the AF—f wave, and lead to false diagnosis. Since the QRS complex is the most prominent feature of an ambulatory ECG and the least sensitive to muscle noise, people have developed different methods for detecting AF by *R–R* intervals, including the coefficient of variation and Kolmogorov–Smirnov test [2], neural networks [3] and Markov models [4]. Logan and Healey used the histogram of variance of *R–R* intervals to identify AF and achieved a sensitivity of 0.96 with a specificity of 0.89 [1]. Tateno and Glass utilized a histogram of the difference between successive *R–R* intervals to detect AF and obtained a sensitivity of 0.932 and specificity of 0.967 [5]. This group later improved the technique [2], achieving a sensitivity of 0.944 and specificity of 0.972. All the methods mentioned above are based on the fact that the irregular *R–R* intervals of AF can be expressed in a typical pattern of *R–R* interval distribution, which could be used to differentiate AF from non-AF rhythms. Nevertheless, it is difficult to detect AF based solely on the *R–R* intervals. The detecting precision still needs to be improved. Besides, for paroxysmal AF, the duration of AF might be as short as 10 s. By using fewer *R–R* intervals, the algorithm has more chances to detect AF. Therefore, the detection response time needs to be shortened.

In this paper, the probability density function (PDF) of the distance between two points in the reconstructed phase space of *R–R* intervals is analyzed. After investigating the different PDF distributions of normal sinus rhythm (NSR) and AF *R–R* intervals, a characteristic parameter k_n was put forward to detect AF. The new algorithm was then tested on the MIT-BIH Atrial Fibrillation database. Next, parameters of the algorithm, such as the number of *R–R* intervals, number

* Corresponding author. Tel.: +86 21 27059203.
E-mail address: hongweilu_jack@hotmail.com (L. Hong-wei).

1350-4533/$ – see front matter © 2008 IPEM. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.medengphy.2008.04.013

of embedding dimensions and slope, were optimized for the best performance.

2. Methods

2.1. Database of signals

The MIT-BIH Atrial Fibrillation database was used to randomly acquire 400 segments of AF and 400 segments of NSR *R–R* intervals from 23 records: 04015, 04043, 04048, 04126, 04746, 04908, 04936, 05091, 05121, 05261, 06426, 06453, 06995, 07162, 07859, 07879, 07910, 08215, 08219, 08378, 08405, 08434 and 08455. There are on average, 18 segments of AF and 18 segments of NSR *R–R* intervals from each record. Each segment has 100 *R–R* intervals. In addition, for long-term *R–R* interval analyses, four AF segments and four NSR segments with 1000 *R–R* intervals were obtained. In record 04126, AF and NSR segments begin at 11:01:10 and 12:28:00, respectively. In record 06995, AF and NSR segments begin at 12:00:00 and 13:33:30, respectively. In record 08405, AF and NSR segments begin at 04:26:00 and 00:00:00. In record 08455, AF and NSR segments begin at 04:24:00 and 00:00:00. We used the automatic *R–R* interval detection algorithm in the Biomedical Signal Processing Laboratory at http://www.bsp.pdx.edu/Toolbox/. The accuracy of the automatic *R–R* interval detection algorithm is about 98% [6]. After manual rectification, the accuracy is almost 100%.

2.2. Phase space reconstruction

The *R–R* interval is just one observable variable from the multi-variable cardiac system. According to Takens' embedding theorem [7], the state of a dynamic system can be represented in a reconstructed phase space by time delay embedding. Taken's theorem guarantees that the dynamic characteristics of the real (physiological) and the reconstructed system are the same. Therefore, it is possible to reconstruct the phase space of the original cardiac system from the *R–R* intervals while preserving the system dynamics. If $x(i)$ is an element of the *R–R* intervals vector $\mathbf{X}$, then a reconstructed vector $\mathbf{Y}_i$ is introduced as

$$\mathbf{Y}_i = [x(i), x(i+\tau), \ldots, x(i+(m-1)\tau)] \quad (1)$$

where τ is the time delay and m is the embedding dimension. The reconstructed vector $\mathbf{Y}_i$ represents a point in the m-dimensional phase space. Phase space reconstruction techniques may provide the physician with useful clinical information based on the cardiac dynamics generating the observed ECG. Thus, in addition to direct interpretation of ECG time series texture, embedding techniques and the subsequent characterization of the ECG in reconstructed phase space may also contribute to the diagnosis of cardiac arrhythmias. By reconstructing the *R–R* intervals and investigating the structure of the reconstructed attractor, it is possible to provide more information and new diagnostic potential of the analyzed cardiac system.

2.3. Probability density function [8]

The basic idea behind the PDF method is to construct the correlation function $C(r,m,\tau)$. $C(r,m,\tau)$ describes the probability that the distance between arbitrary two points $\mathbf{Y}_i$ and $\mathbf{Y}_j$ in the phase space is shorter than distance r. $C(r,m,\tau)$ can be written as

$$C(r, m, \tau) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \theta(r - ||Y_i - Y_j||) \quad (2)$$

where N is the number of phase space points, symbol $||\cdot||$ represents the Euclidean distance, and $\theta(x)$ is the Heaviside unit-step function which is defined by

$$\theta(z) = \begin{cases} 0 & z < 0 \\ 1 & z \geq 0 \end{cases} \quad (3)$$

Eqs. (1)–(3) are essential steps of the Grassberger and Procaccia (GP) method [9] to calculate the correlation dimension. The determination of the correlation dimension from *R–R* intervals are commonly used for gaining information about the nature of the underlying cardiac dynamics [10]. However, the structure of the reconstructed attractor is rarely investigated. What is more, for a reliable determination of the correlation dimension, substantial data (such as 10,000 points) are needed resulting in an enormous computational burden. Since the correlation function $C(r,m,\tau)$ has a maximum of 1, minimum of 0 and is a continuous distribution function, we define the probability density function as

$$p(r, m, \tau) = \frac{dC(r, m, \tau)}{dr} \quad (4)$$

We believe that the probability density function may contain some important information about the spatial distribution of the phase points in the reconstructed attractor. To the best of our knowledge, this is the first time that the PDF method has been proposed for the analysis of the reconstructed attractor structure.

3. Results

3.1. PDF and its distribution

Fig. 1(a) illustrates PDF curves of 1000 NSR *R–R* intervals from record 04126 for time delays of $\tau = 1$, 2, 3, 10, 20 and 30, respectively, where the embedding dimensions m is 5. Fig. 1(b) shows PDF curves of 1000 AF *R–R* intervals from record 04126. Upon curve fitting it is found that for NSR *R–R* intervals, as the time delay gradually increases, the PDF curves gradually change from an asymmetrical to a Gaussian distribution, indicating that the correlation between neighboring NSR *R–R* intervals is strong. For AF *R–R* intervals, PDF