Comparison of different threshold values *r* for approximate entropy: application to investigate the heart rate variability between heart failure and healthy control groups

2011 Physiol. Meas. 32 167

(http://iopscience.iop.org/0967-3334/32/2/002)

IOP PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **32** (2011) 167–180 doi:10.1088/0967-3334/32/2/002

Comparison of different threshold values *r* for approximate entropy: application to investigate the heart rate variability between heart failure and healthy control groups

Chengyu Liu[1], Changchun Liu[1,4], Peng Shao[2], Liping Li[1], Xin Sun[1], Xinpei Wang[1] and Feng Liu[3]

[1] Institute of Biomedical Engineering, School of Control Science and Engineering, Shandong University, Jinan, 250061, People's Republic of China
[2] Department of Electrical and Computer Engineering, University of Alberta, Alberta, T6G2E1, Canada
[3] School of Information Technology and Electrical Engineering, University of Queensland, Queensland, 4072, Australia

E-mail: changchunliu@sdu.edu.cn and bestlcy@mail.sdu.edu.cn

Received 8 September 2010, accepted for publication 24 November 2010
Published 22 December 2010
Online at stacks.iop.org/PM/32/167

Abstract

Approximate entropy (ApEn) is widely accepted as a complexity measure of the heart rate variability (HRV) signal, but selecting the criteria for the threshold value r is controversial. This paper aims to verify whether Chon's method of forecasting the r_{max} is an appropriate one for the HRV signal. The standard limb lead ECG signals of 120 subjects were recorded for 10 min in a supine position. The subjects were divided into two groups: the heart failure (22 females and 38 males, median age 62.4 ± 12.6) and healthy control group (33 females and 27 males, median age 51.5 ± 16.9). Three types of ApEn were calculated: the $\mathrm{ApEn}_{0.2}$ using the recommended constant $r = 0.2$, the ApEn_{chon} using Chon's method and the ApEn_{max} using the true r_{max}. A Wilcoxon rank sum test showed that the $\mathrm{ApEn}_{0.2}$ ($p = 0.267$) and the ApEn_{max} ($p = 0.813$) had no statistical differences between the two groups, while the ApEn_{chon} ($p = 0.040$) had. We generated a synthetic database to study the effect of two influential factors (the signal length N and the ratio of short- and long-term variability sd_1/sd_2) on the empirical formula in Chon's method (Chon *et al* 2009 *IEEE Eng. Med. Biol. Mag.* **28** 18–23). The results showed that the empirical formula proposed by Chon *et al* is a good method for analyzing the random signal, but not an appropriate tool for analyzing nonlinear signals, such as the logistic or HRV signals.

[4] Author to whom any correspondence should be addressed.

Keywords: approximate entropy, heart rate variability, different threshold values r for ApEn

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Approximate entropy (ApEn), which provides a general measure of the regularity of time series, is widely used for analyzing physiological series of clinical data (Pincus 1991, 1995). The popularity of the ApEn stems from its capability in providing quantitative information about the complexity of both short- and long-term data recordings that are often corrupted with noise. The calculation methods are relatively easy and ApEn has been widely applied to clinical cardiovascular studies (Pincus *et al* 1991, Pincus and Keefe 1992, Dawes *et al* 1992, Fleisher *et al* 1993, Ho *et al* 1997, Makikallio *et al* 1998, Pincus 2001).

The ApEn examines the conditional probability of the inherent similarities in the time series based on statistical analyses that appear to be compatible with the general clinical need to distinguish healthy subjects from the abnormal ones (Pincus *et al* 1991). Series that are more repetitive-in-time or with subsequences that are more similar have smaller ApEn values. The calculation of ApEn requires, initially, selection of two unknown parameters: the embedding dimension m and the threshold value r. The parameter m determines the subsequent lengths to be compared; $m = 2$ (Pincus *et al* 1991, Pincus and Keefe 1992) is suggested or it can be estimated by calculating the false nearest neighbor (Kennel *et al* 1992). The second parameter r determines the threshold tolerance for accepting similar patterns between two subsequences and should be in the range between 0.1 and 0.25 times the standard deviation of the time series (Pincus *et al* 1991, Pincus and Keefe 1992). This recommendation was originally used to analyze the relatively slow dynamic time series, such as the heart rate (Fleisher *et al* 1993, Ho *et al* 1997, Hogue *et al* 1998), fetal heart rate (Dawes *et al* 1992, Leeuwen *et al* 2006) and hormonal release data (Pincus *et al* 1999).

Recently, studies have reported that when the dynamic performance of the time series becomes faster, the aforementioned r range may not be appropriate and thus can lead to incorrect conclusions. The r value that maximizes the ApEn comes from Lu *et al* (2008) and Chon *et al* (2009). Hereinafter, we denote the maximum of ApEn as $\mathrm{ApEn}_{\max}$ and the corresponding r value as $r_{\max}$. It is a time-consuming process choosing the $r_{\max}$ from the range 0.01 to 1.0 times the standard deviation of the time series. Instead of the computing the ApEn for every r value, an empirical formula was proposed by Lu *et al* (2008) and Chon *et al* (2009) for automatically selecting the r value to maximize the ApEn. Nevertheless, the empirical formula, derived from the analysis of Gaussian white noise (WN) signals, was based on the method derived from Monte Carlo simulations. Therefore, whether or not it is the most suitable for the analysis of slow nonlinear time series, such as the heart rate variability (HRV) signal, is open to question. Significant differences lie between the Gaussian WN signals and the HRV signals. Castiglioni and Rienzo (2008) verified with data from 10 young healthy volunteers that the $r_{\max}$ calculated by the empirical formula was not incompatible with the traditionally recommended r range and that the choice of r was critical, especially in the HRV studies.

Regardless of how m is chosen—by adopting the recommended value or calculating the false nearest neighbor—m is usually to be chosen as 2 or 3, which slightly affects the analytical results of the ApEn. However, r often imposes an obvious influence on the analytical results

of the ApEn and the influence becomes even more apparent when the dynamics of the time series are uncertain. The purpose of this paper is to verify whether or not the empirical formula for computing $r_{\max}$ given by Chon *et al* (2009) is appropriate for the nonlinear HRV signal analysis. Unlike Castiglioni and Rienzo's 2008 study, we increased the number of subjects from 10 to 120, while adding the comparison between heart failure and healthy control groups, to obtain more accurate conclusions.

2. Methods

2.1. Subjects

The subjects enrolled in this study, either in the heart failure or the healthy control groups, should, (i) be aged between 18 and 75 years old, (ii) agree to sign the information consent form, and (iii) not have participated in any other 'clinical trials' within the previous 3 months.

Individually, the heart failure group must be in accord with classes II–III of the New York Heart Association (NYHA) Functional Classification and have an LVEF<0.50 with an ultrasonic cardiogram (UCG) detection.

The individual requirements for the healthy control group are that (i) half the group should be older than 50 years, and (ii) they should have normal results with a UCG, blood lipid and glucose checks and electrocardiogram (ECG).

Subjects with severe organ damage or subjects with psychiatric disorders will be excluded.

In accordance with the above conditions, 120 subjects were enrolled and were divided into two groups: (i) the heart failure group (22 females and 38 males, median age 62.4 years; range, 38–75 years), and (ii) the healthy control group (33 females and 27 males; median age, 51.5 years; range, 24–72 years). Each group contained 60 subjects. The study had the full approval of the Clinical Ethics Committee of the Qilu Hospitals of Shandong University.

2.2. Data acquisition

All subjects were selected to ensure that they did not take medications or smoke cigarettes before the test. At the beginning, every subject was asked to lie on the bed and relax over a period of time. Subsequently, standard limb II lead ECG data were recorded for about 10 min for each subject, using the cardiovascular system function detecting instrument. The ECG data were sampled at 1000 Hz. Each subject was in the supine position during the recording.

After the data acquisition, the ECG data were filtered through a band-pass filter with its pass-band frequency being set at 0.05–125 Hz. Then, the R-wave peaks of the ECG data were automatically detected by the wavelet transform modulus maxima (WTMM) method (Li *et al* 1995, Martinez *et al* 2004), which was an important method for describing the characteristic elements of a complex quasi-periodic signal based on wavelet transform. The fore-and-aft R-wave peaks formed the R-R interval and the consecutive R-R intervals made up of the original RR sequence.

The original RR sequence often contains two kinds of anomalies: those caused by detector errors and anomalies caused by ectopic beats (Mateo and Laguna 2003). For detector-error caused anomalies, a false beat brought about by a low amplitude R-wave is the usually called a false negative (FN); and the false beat caused by noise masking is a false positive (FP). The anomalies caused by ectopic beats are usually classified into supra-ventricular ectopic beats (sVEB) and ventricular ectopic beats (VEB), depending on the localization of the ectopic focus. Distinct ECG morphological differences exist between these two categories. Usually,

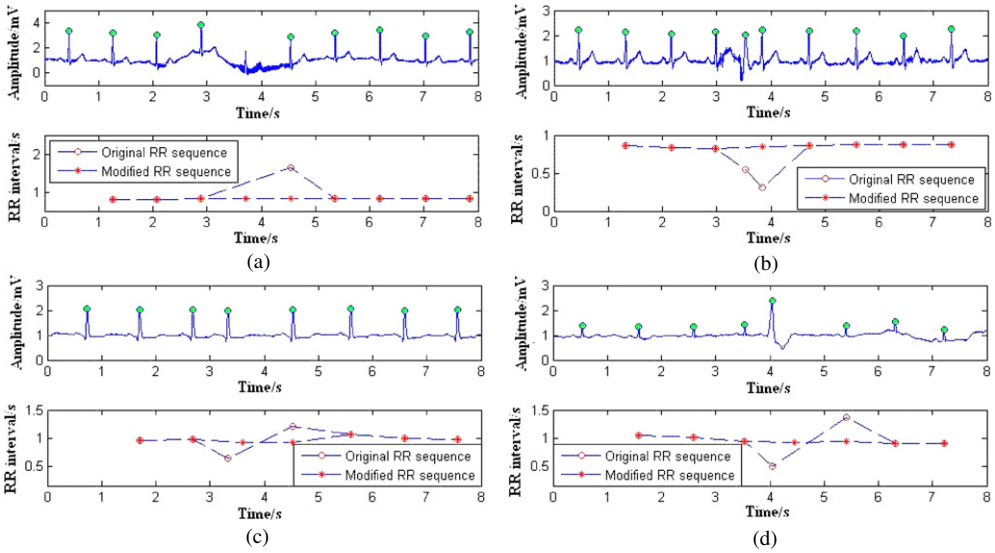

Figure 1. Four different kinds of anomalies in the ECG: (a) FN anomaly, (b) FP anomaly, (c) sVEB anomaly, (d) VEB anomaly.

the morphology features of the sVEB are similar to those of a normal beat, but the VEB has an increased amplitude at the position of the ectopic beat.

Figure 1 shows the four different kinds of anomalies mentioned above: (a) FN anomaly, (b) FP anomaly, (c) sVEB anomaly, and (d) VEB anomaly. Each figure shows a segment of an ECG trace lasting 8 s and the positions of the R-wave peaks automatically detected by the WTMM method in the upper panel. Because these anomalies exhibit a sharp transient in the original RR sequence (see the lower panel of figure 1) that contaminates the real RR sequence, it is necessary to correct the original RR sequence prior to the HRV analysis. In this study, we used the IRF introduced by McNames *et al* (2004) to correct the anomalies in the original RR sequence. In the lower panel of figure 1, the original and modified RR sequences are shown. After correction, sharp transient in the original RR sequence has been removed.

2.3. Chon's method for calculating r_{max}

Recent studies have reported that the $\mathrm{ApEn}_{\mathrm{max}}$ may reflect the complexity of the physiological signal accurately. The $\mathrm{ApEn}_{\mathrm{max}}$ is usually calculated by linearly increasing r within the range 0.01–0.8 with an increment of 0.01. However, it is a time-consuming process when choosing the r_{max}. Based on the analysis of Gaussian white noise (WN) signals, Lu *et al* (2008) and Chon *et al* (2009) proposed an empirical formula for automatically selecting the r_{max}, which is expressed as follows:

$$r_{\mathrm{chon}} = (-0.036 + 0.26\sqrt{sd_1/sd_2})/\sqrt[4]{N/1000}, \tag{1}$$

where N denotes the length of the RR sequence, and sd_1 and sd_2, respectively, are the measure of the short-term and long-term variability of the RR sequence. For an RR sequence, $x(n) = \{x(1), x(2),\ldots, x(N)\}$, let $y(n)$ be the difference sequence of $x(n)$, that is, $y(n) = \{x(2)-x(1),$

$x(3)-x(2),\ldots, x(N)-x(N-1)\}$, then sd_1 is the standard deviation of $y(n)$, and sd_2 is the standard deviation of $x(n)$.

2.4. Definition of different approximate entropies

ApEn was introduced as a quantification of the regularity in the time series, initially motivated by the applications to relatively short, noisy time series. ApEn was derived from the computation of the correlation integral. Developments were made for approximating Markov chains (Pincus 1992). By maximizing the irregularity, it provided a formulation for a finite time series (Pincus and Singer 1996). The elaborate calculation process for the ApEn algorithm can be found from the works of Pincus *et al* (1991) or Hornero *et al* (2005).

In this study, three types of ApEn are discussed according to the different selection criteria of r: $\mathrm{ApEn}_{0.2}$ using the recommended constant $r = 0.2$, $\mathrm{ApEn}_{\mathrm{chon}}$ using the estimated $r_{\max}$ by Chon's method and $\mathrm{ApEn}_{\max}$ using the true $r_{\max}$. The true $r_{\max}$ is found by searching the r values to maximize the ApEn in the range of 0.01–0.8 times the standard deviation of the signal, incrementing by 0.01. Before the calculation of ApEn, each RR sequence of all the 120 subjects is intercepted to be 500 points in length, and the embedding dimension m is set to be 2.

2.5. Statistical analysis

To compare different types of ApEn between the heart failure and healthy control groups, we use the statistical analysis software of SPSS to analyze calculated results. First, we carried out the normal distribution and variance homogeneity test for the indices between the two groups. If positive results were obtained in the normal distribution and variance homogeneity test, we move on to independent sample t-test. If the indices did not pass the tests, we turned to the Wilcoxon rank sum test. $p = 0.05$ was taken as the level of statistical significance for all tests.

3. Results

3.1. Comparison of ApEn and r values calculated using different methods

Four indices of descriptive statistics, the mean, minimum, maximum and standard deviation (SD) are analyzed and shown in table 1. Each measure was separately calculated between the heart failure and healthy control groups. After the normal distribution and variance homogeneity test, the resulting p-values of the Wilcoxon rank sum test are also reported in table 1. A value of $p < 0.05$ is deemed statistically significant. The p-values of $\mathrm{ApEn}_{0.2}$, $\mathrm{ApEn}_{\mathrm{chon}}$ and $\mathrm{ApEn}_{\max}$ are 0.267, 0.040 and 0.813, respectively. While $\mathrm{ApEn}_{0.2}$ and $\mathrm{ApEn}_{\max}$ do not exhibit statistical significances between the two groups, $\mathrm{ApEn}_{\mathrm{chon}}$ behaves significantly differently. The expressive difference of $\mathrm{ApEn}_{\mathrm{chon}}$ between the two groups mainly arose from r_{chon}. Therefore, the estimated r_{chon} can more significantly distinguish the two groups than the $r_{\max}$ in table 1.

Recent studies showed that the maximum value of ApEn (that is, $\mathrm{ApEn}_{\max}$) is likely to behave better than the other values when depicting the complexity of the physiological signals (Chon *et al* 2009, Lu *et al* 2008). Thus, according to this argument, the $\mathrm{ApEn}_{\max}$ in table 1 is supposed to be more representative than the $\mathrm{ApEn}_{\mathrm{chon}}$ and $\mathrm{ApEn}_{0.2}$ in reflecting the inherent complexity of the RR sequence between the heart failure and healthy control groups. Usually, the inherent complexity of the healthy control group is considered higher than that of the heart failure group. However, the results of the Wilcoxon rank sum test show that there is no statistical difference between the two groups in $\mathrm{ApEn}_{\max}$. Similar results also occur with

Table 1. Descriptive statistics of ApEn and r values, the p-values measure the separation between the heart failure and healthy control groups.

	Heart failure group				Healthy control group				
Measures	Mean	Minimum	Maximum	SD	Mean	Minimum	Maximum	SD	p-values
ApEn values									
$\mathrm{ApEn}_{0.2}$	1.085	0.130	1.375	0.203	1.118	0.731	1.302	0.113	0.267
$\mathrm{ApEn}_{\mathrm{chon}}$	1.082	0.093	1.378	0.255	1.156	0.823	1.318	0.105	0.040
$\mathrm{ApEn}_{\mathrm{max}}$	1.174	0.812	1.414	0.138	1.178	0.904	1.332	0.087	0.813
r values									
r_{chon}	0.289	0.131	0.431	0.071	0.253	0.129	0.386	0.056	0.004
r_{max}	0.209	0.026	0.612	0.100	0.231	0.091	0.364	0.057	0.158

Table 2. Descriptive statistics of sd_1, sd_2 and sd_1/sd_2 between the heart failure and healthy control groups.

	Heart failure group				Healthy control group				
Measures	Mean	Minimum	Maximum	SD	Mean	Minimum	Maximum	SD	p-values
sd_1 (ms)	40.634	1.992	346.940	67.125	29.871	5.002	29.804	23.701	0.254
sd_2 (ms)	36.500	1.616	242.632	46.069	36.041	8.335	36.001	19.655	0.938
sd_1/sd_2	0.947	0.329	1.838	0.380	0.799	0.263	1.587	0.283	0.017

$\mathrm{ApEn}_{0.2}$. Therefore, $\mathrm{ApEn}_{\mathrm{max}}$ and $\mathrm{ApEn}_{0.2}$ do not approve this argumentation. Surprisingly, $\mathrm{ApEn}_{\mathrm{chon}}$ is statistically different between the two groups. The reason should to be investigated and the key might lie in the empirical formula (1).

3.2. Analysis of sd_1 and sd_2 for two groups

For each RR sequence of all 120 subjects, we calculate the short-term variability sd_1 and long-term variability sd_2 used in the empirical formula (1). The RR sequence is set to be the uniform length $N = 500$. The statistic results are shown in table 2. Compared with the healthy control group, the heart failure group has a similar sd_2 but a fairly large sd_1. Therefore, sd_1/sd_2 becomes larger in the heart failure group than in the other group. The boxplots of sd_1, sd_2 and sd_1/sd_2 are shown in figure 2. In figure 2, we can clearly discern that the distribution range of sd_1 in the heart failure group varies more intensively than that in the healthy control group. The psychopathological explanation is that the course of heart failure is often accompanied by fatal arrhythmias, such as supra-ventricular ectopic beats, ventricular ectopic beats, and even ventricular fibrillation or atrial fibrillation. This can cause the acute fluctuations in RR sequence. So sd_1 will become large. Sometimes, the RR sequence of the heart failure subject without arrhythmia has a small range due to the weakening of regulatory functions of the autonomic nervous system. So sd_1 will become low. The huge range of sd_1 is the essential reason why sd_1/sd_2 exhibits a significant difference between the two groups and the resulting significant difference in r_{chon} and $\mathrm{ApEn}_{\mathrm{chon}}$. In the next section, we will discuss the influential factors for the empirical formula (1), that is, we explore the influential factors for r_{chon}.

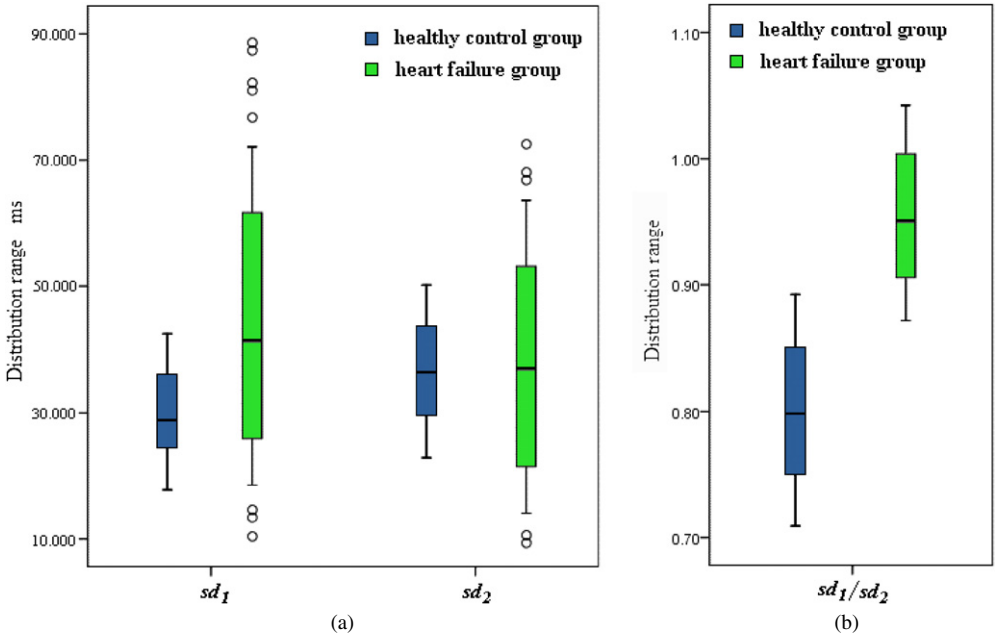

Figure 2. The boxplots of sd_1, sd_2 and sd_1/sd_2 between heart failure and healthy control groups: (a) sd_1 and sd_2, (b) sd_1/sd_2.

3.3. Influential factors analysis for r_{chon}

3.3.1. Construction of the emulation database. In the empirical formula (1), there are two influential factors, N, sd_1/sd_2. To determine whether or not each of the factors separately affects r_{chon}, we carried out experiments with an emulation database composed of random sequences and logistic sequences. The random sequences were generated by the normally distributed random number with the standard deviation 1 and named from R1 to R100. Each of the random sequences has a length of 1000. To eliminate random factors, the ApEn is the average of all ApEn values from 100 sequences. The logistic sequence is considered an approved nonlinear chaotic sequence. In the present project, we use sequences generated by the following iteration function:

$$x(n+1) = \omega * x(n) * (1 - x(n)), \tag{2}$$

where the initial value $x(0)$ is in the range from 0.1 to 0.9. ω, and is a constant parameter that determines the complexity of the sequence. As ω increases, the complexity of the sequence increases. Herein, ω took the value of 3.6, 3.8 and 3.9, and the corresponding logistic sequences were recorded as L1, L2 and L3. To eliminate the effect of the random factors, 90 samples were produced with L1, L2 and L3, and the mean of each sample was used to compute the ApEn. These 90 samples were produced as follows: $x(0)$ in formula (2) was taken as 0.1, 0.2, . . ., 0.9 in turn, with an increment of 0.1. Corresponding to each $x(0)$, a logistic sequence with the length $N = 54\,000$ was produced, and then the logistic sequence was divided up into nine sequence fragments with a length of $N = 6000$. The last 1000 points of each sequence fragment was chosen as a sample. Thus, 90 samples were formed in total.

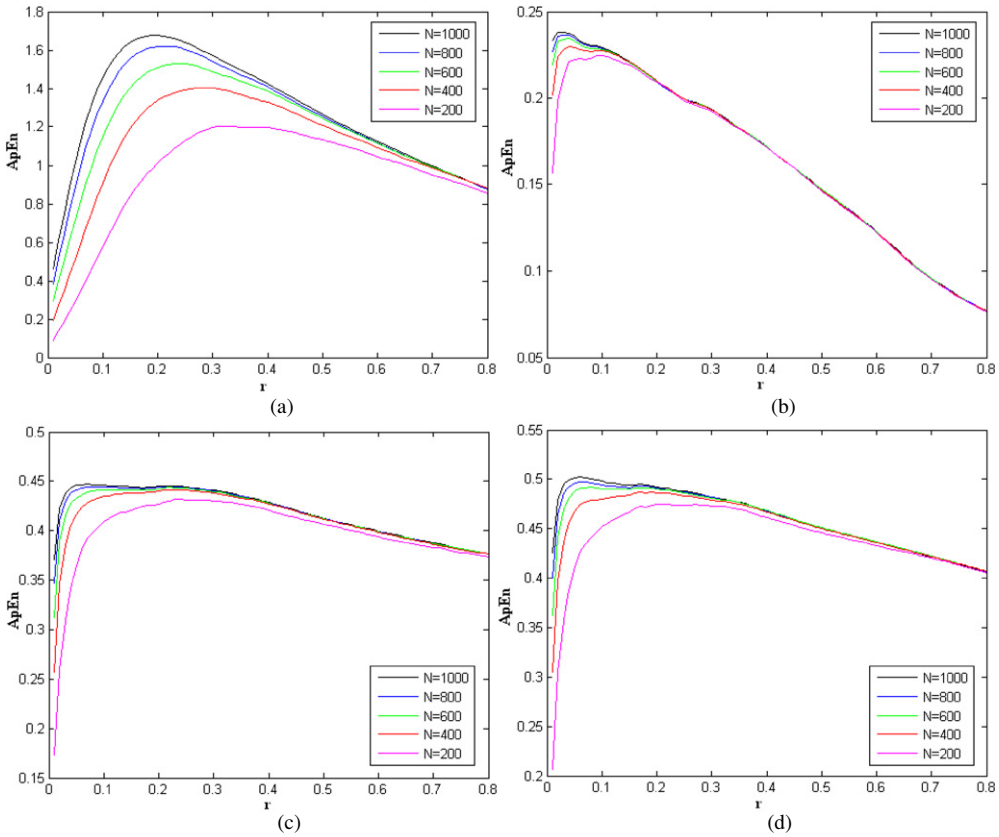

Figure 3. The trend of ApEn with r, including the random sequence and logistic sequence: (a) random sequence, (b) L1 sequence, (c) L2 sequence, and (d) L3 sequence.

3.3.2. How N affects r_{chon}. Figure 3 shows the trend of ApEn with r when N varies. There are four sequences, including the random sequence, L1, L2 and L3. The original sequence has a length of 1000 and, on this basis, is truncated to 800, 600, 400, and 200 points, respectively. The range of r is 0.01–0.8 times the standard deviation of the sequence, with an increment of 0.01. The r corresponding to the $\mathrm{ApEn}_{\max}$ is defined as $r_{\max}$. The contrasted of random and logistic sequence can be found when the variation of $r_{\max}$ is different from the variation of the sequence length N. The $r_{\max}$ in the random sequence has a regular variation with the change of N. If N becomes larger, the corresponding $r_{\max}$ is smaller. This phenomenon is consistent with the empirical formula (1). However, for the logistic sequence, $r_{\max}$ does not show the orderliness and $\mathrm{ApEn}_{\max}$ does not show a distinct single peak as a random sequence. $\mathrm{ApEn}_{\max}$ in the logistic sequence maintains a largish value in a wide range of r, especially for L2 and L3. It is very difficult and impractical to forecast the $r_{\max}$ using the empirical formula.

We assume that the $sd1/sd2$ in the empirical formula (1) is a constant 1.5; thus the empirical formula (1) becomes a single-variable function reflecting the relation between $r_{\max}$ (that is, r_{chon}) and N. The function curve of $r_{\max}(N)$ forecasted by the empirical formula (1) is shown in figure 4 with the black curve. Figure 4 also shows the $r_{\max}(N)$ of the sequences. Because the curves of ApEn (r) in figure 3 are not very smooth, it is not convincing to select

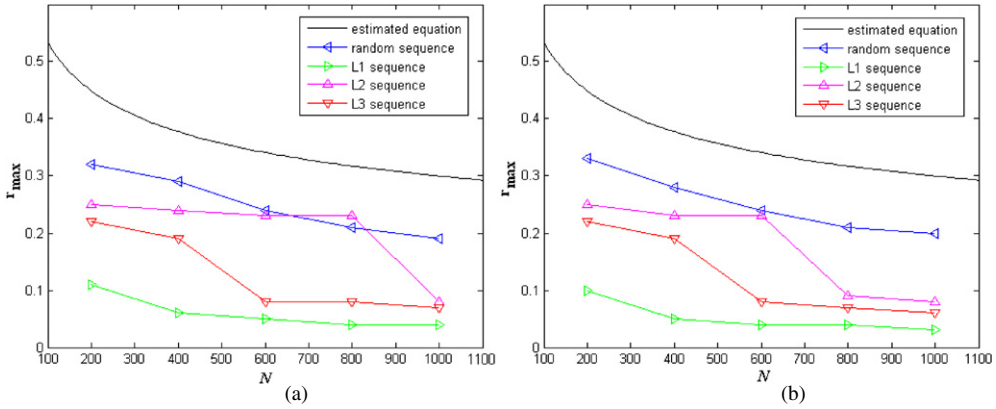

Figure 4. The variable relation of r_{max} and N in the empirical formula (1) and the emulation database; r_{max} is calculated by setting the glide-window p. (a) $p = 3$. (b) $p = 5$.

the r corresponding to the maximum of the ApEn(r) as the r_{max}. We use a glide-mean method to search r_{max} and build a function as follows:

$$\mathrm{ApEn}_{max}(r) = \sum \mathrm{ApEn}(r + 0.01 * m) \qquad |m| \leqslant (p-1)/2, \tag{3}$$

where p is the length of the glide window, r is in the range of 0.01*(p/2+0.5) and 0.8–0.01 *(p/2–0.5). The r maximizing the $\mathrm{ApEn}_{max}(r)$ is defined as r_{max}. The results corresponding to $p = 3$ and $p = 5$ are shown in figures 4(a) and (b), respectively. With the increase of the sequence length N, the r_{max} of the random sequence maintains a consistent trend with the empirical formula (1), but for the logistic sequence, especially when the sequence complexity increases ($\omega = 3.8$ and $\omega = 3.9$), the r_{max} does not show the trend in the empirical formula (1). In addition, as the glide-window p changes, the r_{max} shows some saltation even though with the same N, which indicates that the variable relation of r_{max} and N in the empirical formula (1) is open to question.

3.3.3. How sd_1 and sd_2 affect r_{chon}. Figure 5 shows the boxplots of sd_1 and sd_2 in random and L3 sequences, in which the random sequences include 100 samples and the L3 sequences include 90 samples. It can be found that in either the random or L3 sequences, with sequence length N increasing, the fluctuant ranges of sd_1 and sd_2 become smaller and behaves more stable. Besides, compared with the other N, the fluctuant ranges of sd_1 and sd_2 reaches maximum with $N = 200$, indicating that the stability using the sequence with $N = 200$ to predict the r_{max} is fallacious and the ApEn_{max} on this basis has little credibility.

To discuss the single-variable relationship between the r_{max} and sd_1/sd_2, we set N a fixed number of 600. We add each random and L3 sequence with a low-frequency sine wave sharing the same oscillation period but different amplitudes. The oscillation period is 100 points per cycle. The oscillation amplitude is set to be A times the magnitude of $x(n)$. A is set to be 0, 0.1, 0.2, 0.3, 0.4 and 0.5, respectively. Figure 6 shows two samples respectively from the random and L3 sequences, on which low-frequency sine waves added. A is respectively 0, 0.1, 0.2, 0.3, 0.4 and 0.5 from top to bottom. Because of the low-frequency nature of the oscillation, the short-term variability sd_1 has only slight variation, whereas the long-term variability sd_2 changes more significantly. This is shown in figure 7. The sd_1 remains essentially unchanged,

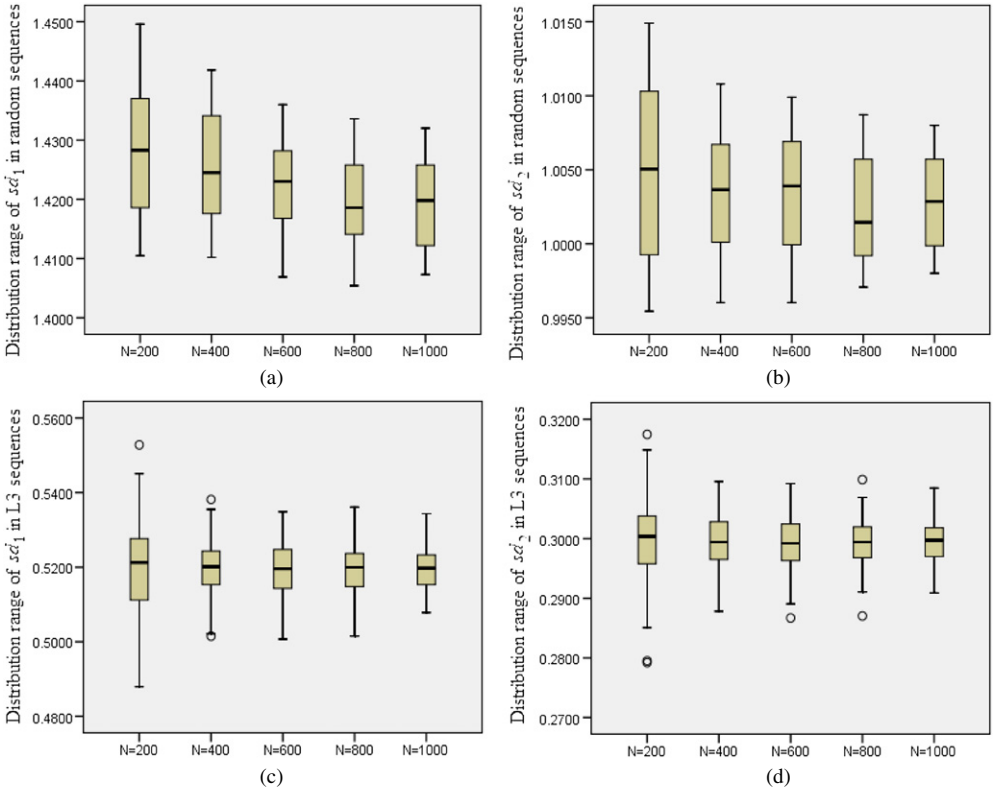

Figure 5. The boxplots of sd_1 and sd_2 in the random and L3 sequences: (a) sd_1 in random sequences, (b) sd_2 in random sequences, (c) sd_1 in L3 sequences, (d) sd_2 in L3 sequences.

but with the increase of oscillation amplitude in the sine wave, the fluctuation range of sequence becomes larger, so the sd_2 significantly increases.

Based on the above analysis, the sd_1/sd_2 will change when A increases. It is reasonable that the $r_{\max}$ will also change according to the empirical formula (1). We calculate the different $r_{\max}$ values in the random and L3 sequences corresponding to the different sd_1/sd_2. Each $r_{\max}$ in the random and L3 sequences is calculated with the average of all the samples. The trends of $r_{\max}$ with sd_1/sd_2 increasing are shown in figure 8. For comparison, we also depict the sd_1/sd_2–$r_{\max}$ curve by the empirical formula (1). Herein, N is set to 600. Similar to section 3.3.2, we also use the glide-mean method to search $r_{\max}$. The glide-window p also be set to $p = 3$ and $p = 5$.

Figure 8 shows that the results of the random sequence are coincident with the empirical formula (1); but for the nonlinear sequence, the test and prediction results differ observably. The trends of the test and prediction results are contrary rather than being consistent. With the sd_1/sd_2 increasing, $r_{\max}$ decreased in the test, but by contrast, it is supposed to increase according to the empirical formula (1). This proves once again that using the empirical formula (1) to forecast the $r_{\max}$ for nonlinear sequence is not accurate, as aforementioned. Because similar to the logistic sequences, the physiological signals, such as the HRV signal, are generally recognized as nonlinear signals.

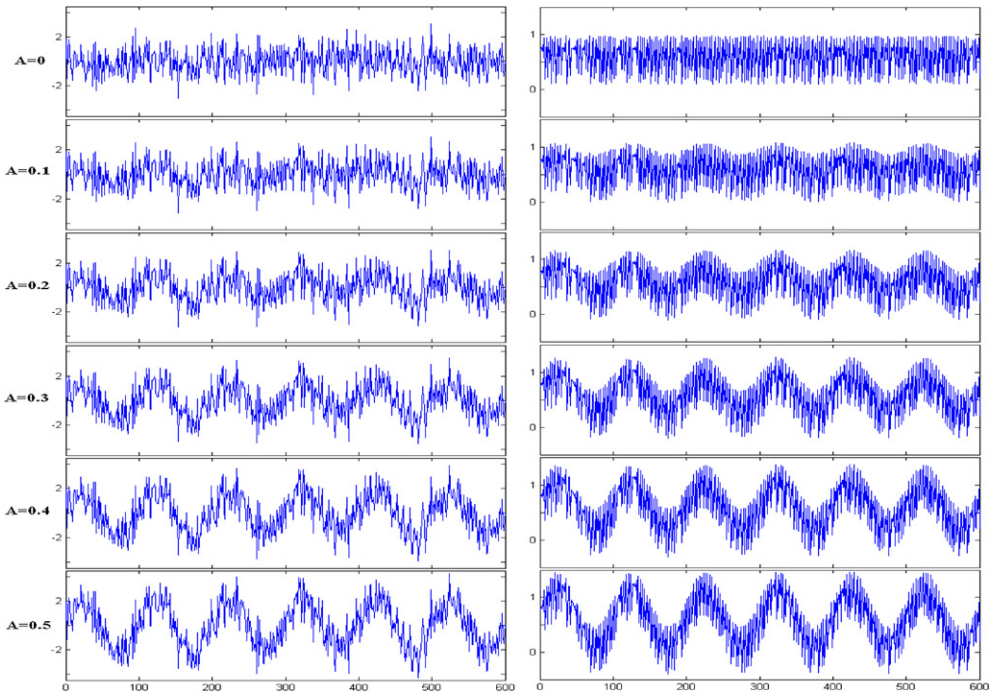

Figure 6. Illustrational samples from the random and L3 sequences added with a low-frequency sine wave: (a) random sequences, (b) L3 sequences.

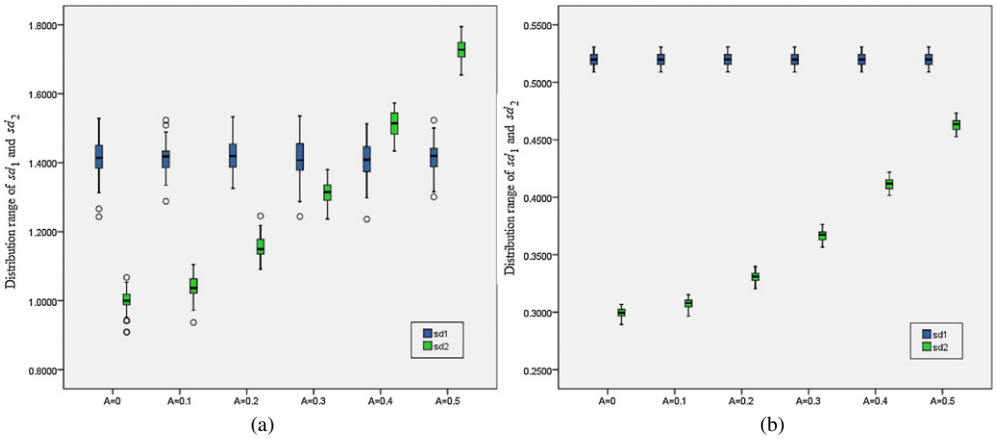

Figure 7. The clustered boxplots of sd_1 and sd_2 in random and L3 sequences: (a) random sequences, (b) L3 sequences.

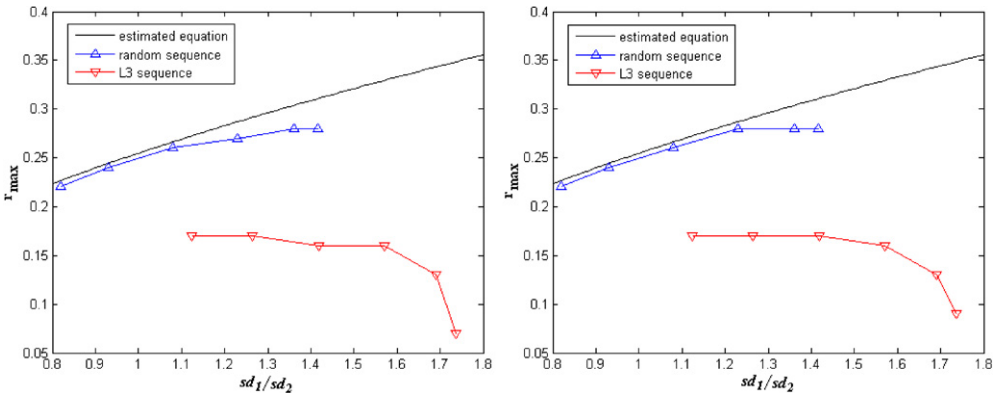

Figure 8. The variable relation of r_{max} and sd_1/sd_2 in estimated equation (1) and emulation database; r_{max} is calculated by the formula (3) by setting the glide-window p. (a) $p = 3$. (b) $p = 5$.

4. Discussion

Since Pincus proposed the ApEn algorithm in 1991, ApEn has been applied to many physiological signal analyses as a measure of signal complexity. The typical applications include the heart rate (Fleisher *et al* 1993, Makikallio *et al* 1998, Gonçalves *et al* 2008), fetal heart rate (Dawes *et al* 1992, Leeuwen *et al* 2006, Gonçalves *et al* 2007), hormone pulsatility (Pincus and Keefe 1992, Pincus *et al* 1999), EEG (Abásolo *et al* 2005, Bruhn *et al* 2000). However, the validity of ApEn is clearly affected by the pre-established threshold value r. Over the past ten years, researchers usually directly use the recommended r in the range between 0.1 and 0.25 times of the standard deviation of the time series. Recently, some researchers point out that the recommended r does not fit all situations and may lead to the wrong results, especially when the signal becomes fast. The use of the r value that maximizes the ApEn ($ApEn_{max}$) is proposed instead of taking a fixed r. The hypothesis is that the $ApEn_{max}$ is able to reflect the true complexity of the different physiological signals more clearly. Unfortunately, the procedure to search for r_{max} is computationally expensive. Chon *et al* (2009) propose an empirical formula to forecast the r_{max} based on the statistical analysis of random signals. Their work makes that it is an easy method to obtain the r_{max} using the empirical formula (1) and then to calculate the $ApEn_{max}$ ($ApEn_{chon}$).

In this study, we intend to investigate whether Chon's method is appropriate for HRV signal analysis. One hundred twenty subjects were enrolled and were divided into two groups: the heart failure and healthy control groups. We calculated three types of ApEn: $ApEn_{0.2}$ using the recommended constant $r = 0.2$, $ApEn_{chon}$ using Chon's method and $ApEn_{max}$ using the r_{max}. Then we analyzed the statistical difference of different ApEn values between the heart failure and healthy control groups using the Wilcoxon rank sum test. We also analyzed the statistical difference of the estimated r_{max} (r_{chon}) and true r_{max}. The Wilcoxon rank sum test showed that $ApEn_{0.2}$ ($p = 0.267$) and $ApEn_{max}$ ($p = 0.813$) had no statistical difference between the two groups, while $ApEn_{chon}$ ($p = 0.040$) had. Similar situations occurred in the r_{chon} ($p = 0.004$) and r_{max} ($p = 0.158$). If $ApEn_{max}$ reflects the true complexity of the physiological signals more accurately, the statistical difference of $ApEn_{max}$ between the heart failure and healthy control groups should be more significant than that of any other ApEn measures. But the results do not support this viewpoint. The secret is likely to be concealed

in the differences between r_{chon} and r_{max}. Thus, the process of calculating the r_{chon} should to be explored, and the validity using the empirical formula to forecast the r_{max} is dubitable.

We used the emulation database to analyze two influence factors for r_{chon}, the signal length N and the ratio of short- and long-term variability, sd_1/sd_2. Disparate results were obtained from the random and logistic sequence. When we fixed the sd_1/sd_2 as a constant, we obtained the conclusion that with N increasing, the true r_{max} was almost like the estimated r_{chon} in the random sequence but not in the logistic sequence. The r_{max} in the logistic sequence did not justify the orderliness forecasted by the empirical formula (1). When N was fixed as a constant and sd_1/sd_2 increased, the trends of true r_{max} were found well coincident with the empirical formula in the random sequence, but not in the logistic sequence. The r_{max} in the logistic sequence that exhibited an essentially disaccord with the estimated r_{chon}, even became contrary. In conclusion, prudence should be exercised when estimating r_{max} using the empirical formula. The reason maybe was that Chon's empirical formula was built on the analysis of the random signal, and the HRV signal was essentially a nonlinear signal. Thus, the validity of the empirical formula to analyze the HRV signal needed to be discussed.

It has been confirmed that HRV analysis has an important effect for the early detection and quantitative evaluation of cardiovascular diseases. The accurate analysis methods for HRV are the premise to the effective clinical practices. This paper carefully compared different criteria proposed by other researchers for selecting threshold values r for ApEn. ApEn is advised to be used with caution because different threshold values r could affect the analysis results, and even educe the opposite conclusion. To search for methods to weaken the influence from the different threshold values r in ApEn computation is among the list of our future work.

Acknowledgments

This work was supported by the Hi-Tech Research and Development Program of China (no 2006AA02Z4D9 and 2009AA02Z408). The authors would like to thank Professor LL and her colleagues for their support in the data collection process at the Qilu Hospital of Shandong University, Jinan, China.

References

Abásolo D, Hornero R, Espino P, Poza J, Sánchez C I and Rosa R 2005 Analysis of regularity in the EEG background activity of Alzheimer's disease patients with approximate, entropy *Clin. Neurophysiol.* **116** 1826–34

Bruhn J, Röpcke H, Rehberg B, Bouillon T and Hoeft A 2000 Electroencephalogram approximate entropy correctly classifies the occurrence of burst suppression pattern as increasing anesthetic drug effect *Anesthesiology* **93** 981–5

Castiglioni P and Rienzo M D 2008 How the threshold 'r' influences approximate entropy analysis of heart-rate variability *Comput. Cardiol.* **35** 561–4

Chon K H, Scully C G and Lu S 2009 Approximate entropy for all signals *IEEE Eng. Med. Biol. Mag.* **28** 18–23

Dawes G S, Moulden M, Sheil O and Redman C W 1992 Approximate entropy, a statistic of regularity, applied to fetal heart rate data before and during labor *Obstet. Gynecol.* **80** 763–8

Fleisher L A, Pincus S M and Rosenbaum S H 1993 Approximate entropy of heart rate as a correlate of postoperative ventricular dysfunction *Anesthesiology* **78** 683–92

Gonçalves H, Bernardes J, Rocha A P and Ayres-de-Campos D 2007 Linear and nonlinear analysis of heart rate patterns associated with fetal behavioral states in the antepartum period *Early Hum. Dev.* **83** 585–91

Gonçalves H, Henriques-Coelho T, Bernardes J, Rocha A P, Nogueira A and Leite-Moreira A 2008 Linear and nonlinear heart-rate analysis in a rat model of acute anoxia *Physiol. Meas.* **29** 1133–43

Ho K K, Moody G B, Peng C K, Mietus J E, Larson M G, Levy D and Goldberger A L 1997 Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics *Circulation* **96** 842–8

Hogue C W Jr, Domitrovich P P, Stein P K, Despotis G D, Lisa R, Schuessler R B, Kleiger R E and Rottman J N 1998 RR interval dynamics before atrial fibrillation in patients after coronary artery bypass graft surgery *Circulation* **98** 429–34

Hornero R, Aboy M, Abasolo D, McNames J and Goldstein B 2005 Interpretation of approximate entropy: analysis of intracranial pressure approximate entropy during acute intracranial hypertension *IEEE Trans. Biomed. Eng.* **52** 1671–80

Kennel M B, Brown R and Abarbanel H D 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction *Phys. Rev.* A **45** 3403–11

Leeuwen P V, Cysarz D, Lange S and Gronemeyer D 2006 Increase in regularity of fetal heart rate variability with age *Biomed. Tech. (Berl.)* **51** 244–7

Li C, Zheng C and Tai C 1995 Detection of ECG characteristic points using wavelet transform *IEEE Trans. Biomed. Eng.* **42** 21–8

Lu S, Chen X, Kanters J K, Solomon I C and Chon K H 2008 Automatic selection of the threshold value R for approximate entropy *IEEE Trans. Biomed. Eng.* **55** 1966–72

Makikallio T H, Ristimae T, Airaksinen K E, Peng C K, Goldberger A L and Huikuri H V 1998 Heart rate dynamics in patients with stable angina pectoris and utility of fractal and complexity measures *Am. J. Cardiol.* **81** 27–31

Martinez J P, Almeida R, Olmos S, Rocha A P and Laguna P 2004 A wavelet-based ECG delineator: evaluation on standard databases *IEEE Trans. Biomed. Eng.* **51** 570–81

Mateo J and Laguna P 2003 Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal *IEEE Trans. Biomed. Eng.* **50** 334–43

McNames J, Thong T and Aboy M 2004 Impulse rejection filter for artifact removal in spectral analysis of biomedical signals *Proc. 26th Annu. Int. Conf. of the IEEE EMBS (San Francisco, CA, USA)* pp 145–8

Pincus S 1992 Approximating Markov chains *Proc. Natl Acad. Sci. USA* **89** 4432–6

Pincus S and Singer B 1996 Randomness and degrees of irregularity *Proc. Natl Acad. Sci. USA* **93** 2083–8

Pincus S M 1991 Approximate entropy as a measure of system complexity *Proc. Natl. Acad. Sci. USA* **88** 2297–301

Pincus S M 1995 Approximate entropy (ApEn) as a complexity measure *Chaos* **5** 110–7

Pincus S M 2001 Assessing serial irregularity and its implications for health *Ann. New York Acad. Sci.* **954** 245–67

Pincus S M, Gladstone I M and Ehrenkranz R A 1991 A regularity statistic for medical data analysis *J. Clin. Monit.* **7** 335–45

Pincus S M, Hartman M L, Roelfsema F, Thorner M O and Veldhuis J D 1999 Hormone pulsatility discrimination via coarse and short time sampling *Am. J. Physiol.* **277** 948–57

Pincus S M and Keefe D L 1992 Quantification of hormone pulsatility via an approximate entropy algorithm *Am. J. Physiol.* **262** 741–54