Dynamical complexity detection in short-term physiological series using base-scale entropy

Jin Li[1,2] and Xinbao Ning[1,*]

[1]*State Key Laboratory of Modern Acoustics, Institute for Biomedical Electronical Engineering, Nanjing University, Nanjing 210093, China*

[2]*College of Physics and Information Technology, Shaanxi Normal University, Xi'an 710062, China*

(Received 19 May 2005; revised manuscript received 15 December 2005; published 24 May 2006)

Physiological systems generate complex fluctuations in their output signals that reflect the underlying dynamics. The base-scale entropy method was proposed as a complexity measure to investigate the complexity of time series. The advantages of this method are simplicity and extremely fast calculation for very short data sets. This method enables analyzing very short, nonstationary, and noisy data sets. We employed this method for short-term physiological time series for analysis of heart-rate variability signals. The results show that the simple and easily calculated measure can effectively detect the complexity dissimilarity of physiological time series in different physiological or pathological states, which is convenient for clinical applications.

DOI: 10.1103/PhysRevE.73.052902 PACS number(s): 87.10.+e, 87.19.Hh, 05.45.Tp, 02.50.−r

I. INTRODUCTION

Dynamic complexity detection for outputting time series of complex systems is one of the most important problems in physics, biomedicine, engineering, and economic sciences. Especially in biomedicine, accurate detection of the dissimilarity between normal and abnormal states may improve diagnosis and treatment and also provide convenience for clinical applications.

Many ingenious algorithms and methods have been developed during the last 20 years in order to estimate complexity measures from real-world time series, such as dimensions and Lypunov exponents [1–5]. They are both working well, but they generally require long data sets for statistically significant results, which results in inconvenience in clinical studies and applications. Another main type of complexity measures is the entropy. Entropy has the advantages of simplicity, extremely fast calculation, and antinoise ability. They provide convenience for detecting and capturing useful information of time series [6–13]. Some entropy methods based on symbolic dynamics adopt a range partition to generate a partition in the symbolization transform, but these meaningful results may be compromised by the nonstationarity of the time series [14]. Steuer *et al.* [15] reported that a partition according to the principle of maximized entropy gave a better tool to differentiate sequences than the usually used homogeneous partition. For the well-known chaotic dynamic systems, logistic maps, the T entropy values introduced by Ebeling *et al.* [16] approach the Lyapunov exponent asymptotically for increasing string length. From their work we got the idea that the entropy measure should play an important role if we use a much more reasonable partition method. The data sets obtained from most clinical and physiological studies usually are nonstationary, rather short, and noisy, such as heart-rate variability (HRV) signals. Our objective was to find an effective method that requires very short data sets for statistically significant results, provides the ability to make fast calculations, and can be used to analyze nonstationary and noisy data, which is convenient for the analysis of real-world time series.

II. THEORY

We consider a time series u of N points as follows: $\{u(i), 1 \leqslant i \leqslant N\}$. First we embed the time series in an m-dimensional space [17]. For every point $u(i)$, we selected m data points from the series to make an m-dimensional vector

$$X(i) = [u(i), u(i+L), \ldots, u(i+(m-1)L)], \tag{1}$$

where m is the embedding dimension and L the delay time. Choosing $L=1$, the number of m-dimensional vectors is $N-m+1$ in this paper. For each m-dimensional vector, the base scale (Z_{BS}) is calculated by defining the base scale as the root mean square of the differences between every two contiguous data points in an m-dimensional vector,

$$Z_{BS}(i) = \sqrt{\sum_{j=1}^{m-1} [u(i+j) - u(i+j-1)]^2/(m-1)}. \tag{2}$$

Based on the base scale, the partition standard can be selected as $a \times Z_{BS}$. We transformed each m-dimensional vector $X(i)$ into a symbolic sequence $S_i(X(i)) = \{s(i), s(i+1), \ldots, s(i+m-1)\}$, $s \in A$, on the basis of the alphabet A ($A=0,1,2,3$). The transformation into symbols refers to four given levels:

$$S_i(X(i)) = \begin{cases} \bar{u} < u_{i+k} \leqslant \bar{u} + a \times Z_{BS}, & 0, \\ u_{i+k} > \bar{u} + a \times Z_{BS}, & 1, \\ \bar{u} - a \times Z_{BS} < u_{i+k} \leqslant \bar{u}, & 2, \\ u_{i+k} \leqslant \bar{u} - a \times Z_{BS}, & 3, \end{cases} \tag{3}$$

where $i=1,2,3,\ldots,N-m+1$, $k=0,1,2,\ldots,m-1$. $\bar{u}$ denotes the mean of the ith m-dimensional vector $X(i)$, and Z_{BS} denotes the base scale of the ith m-dimensional vector $X(i)$. The symbols 0, 1, 2, and 3 are only used as a mark for every

*Email address: xbning@nju.edu.cn

region; the values have no practical meaning. a is a special parameter. If a is too large during the course of transforming the original time series into symbolic time series, the detailed information will be lost and the dynamic information cannot be captured. Too small of a value of a will result in salient influence from noise. In this paper, we obtain the a value by testing the method used by Wessel *et al.* [13].

We studied the probability distribution about the symbolic sequences S_i of the m-dimensional vector $X(i)$ in order to calculate the base-scale entropy. The symbolic sequences S_i (m-words) at most have 4^m different forms π since they are made up of four symbols, 0, 1, 2, and 3. For each π, we determine the relative frequency

$$p(\pi) = \text{No.}\{t|1 \leqslant t \leqslant N-m+1 \ (u_t, \ldots, u_{t+m-1}) \text{ has form } \pi\}/(N-m+1). \quad (4)$$

The base-scale entropy of the m-dimensional vector is defined as

$$H(m) = -\sum P(\pi)\log_2 P(\pi), \quad (5)$$

where the value of $m=3$, 4, 5, 6, 7, which seems to be suitable for calculating convenience, and it is also reasonable that N is larger than 4^m. The sum is for all forms π. This is the wave information contained in m consecutive values of the time series. It is clear that $0 \leqslant H(m) \leqslant \log_2 4^m$, where the lower bound is attained for a series that presents only one form among 4^m possible forms, and the upper bound for a completely random series where all 4^m possible forms appear with the same probability. The time series presents many types of dynamic forms when $H(m) < \log_2 4^m$.

The base-scale entropy method essentially quantifies the uncertainty of the occurrence of m-words form π. The larger entropy denotes the more uncertain occurrence of m-words form π. We use the word "complexity" to denote the uncertainty. The greater complexity denotes the larger uncertainty, and vice versa. This method is completely different from other symbol dynamic methods in the transformation of symbols. Since the base scale of each m-dimensional vector is different, the standard of partitioning is also changed dynamically. We used a dynamical adaptive partitioning approach in transforming the time series into a symbol series. The base-scale entropy method proposes using the base scale in symbol transformation as a renormalization of the delay vectors, which leads to an equal coding form of similar wave forms by rescaling the amplitude as illustrated in Fig. 1. The purpose is to extract the wave characteristics of the time series and ignore amplitude information. In Fig. 1, $X(i)$ and $X(j)$ are five-dimensional vectors. Their wave modes are similar. Transformed into symbols, the two m-words are all 23131 though they have different wave amplitude and mean. It is possible using this method to analyze nonstationary time series. To further explain this nomenclature, we discuss the relation between the base scale and standard deviation. From Eq. (2)

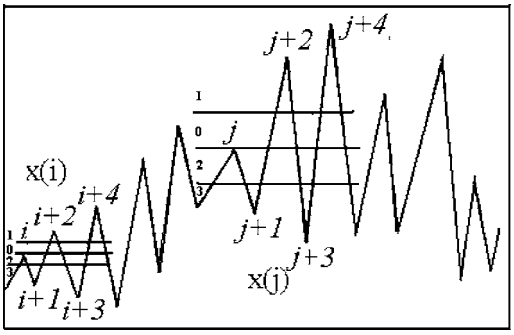

FIG. 1. Illustration of similar wave mode.

$$Z_{BS}(i) = \sqrt{\sum_{j=1}^{m-1} [u(i+j) - u(i+j-1)]^2/(m-1)} = \sqrt{E(u_{i+1} - u_i)^2}.$$

When the series is a stochastic sequence and its mean is 0, the $Z_{BS}(i)$ is given by

$$Z_{BS}(i)|_{stochastic} = \sqrt{E(u_{i+1} - u_i)^2} = \sqrt{E(u_{i+1}^2 + u_i^2 - 2u_{i+1}u_i)} = \sqrt{E(u_{i+1}^2) + E(u_i^2)} = \sqrt{2} \times \sigma_i|_{stochastic}, \quad (6)$$

σ is the standard deviation of the series. We can see that the base scale changes with the standard deviation. A statistical property such as mean or standard deviation does not remain the same throughout the time recording for a nonstationary time series. The base-scale entropy method, which used a dynamic adaptive partition to investigate the complexity of time series, can also be applied to analyze nonstationary data.

III. BASE-SCALE ENTROPY APPLIED TO MODEL SYSTEMS

Using a logistic map given by $x_{i+1} = r x_i(1-x_i)$, where r is the control parameter, we generated the time series $x(r)$ of the logistic map by starting from $x_0 = 0.65$, $r_0 = 3.4$ and continuously increased r in steps of 10^{-4}. Figure 2(a) shows the resulting time series. The vertical dashed lines indicate the bifurcations and periodic windows. Figure 2(b) shows the base-scale entropy of the logistic map. We used $m=6$, $N=9000$, $a=2.0$ for calculating the entropy value. We tested several values of a from 0.1 to 2, and the results showed no significant difference. The parameter $m=4,5,6,7$ also had little influence on the results. Figure 2(c) shows the Lyapunov exponent of the logistic map. It is clear that Figs. 2(b) and 2(c) have a very similar appearance in the whole chaotic regions and both show the transformation of dynamic complexity very well. This result also resembles that obtained with the permutation entropy [18]. The different appearance within the periodic region $r < 3.57$ shows that the base-scale entropy quantifies the periodic series better than the Lyapunov exponent does.

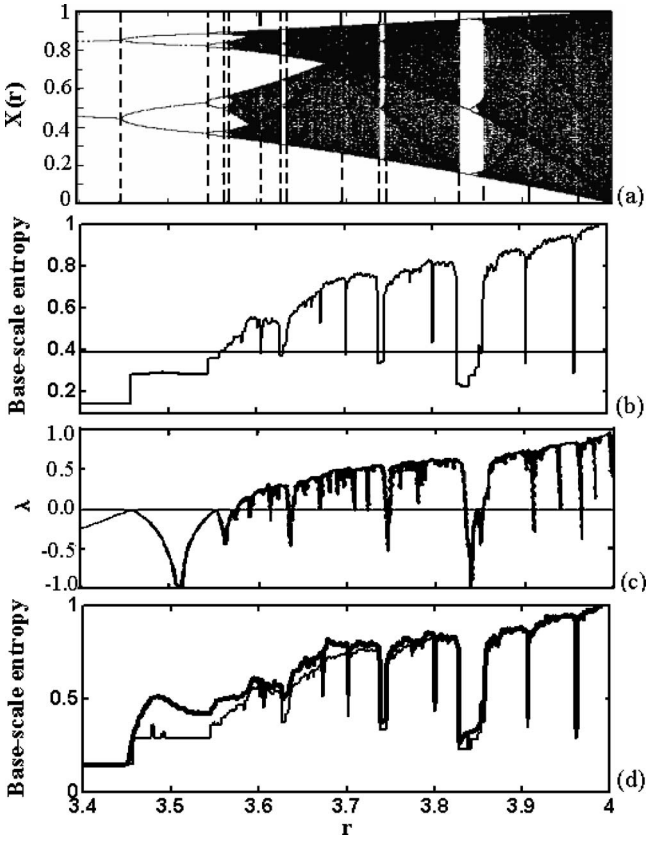

FIG. 2. (a) Logistic map for varying parameter r. (b) The normalized base-scale entropy for the logistic map. (c) The Lyapunov exponent λ for the logistic map. (d) The normalized base-scale entropy with Gaussian noise at standard deviations $s=0.000\,25$ (thin line) and 0.1 (thick line).

The base-scale entropy method is effective in the presence of noise. In theory, to avoid a significant influence from noise in a base-scale entropy calculation, the partition standard $a \times Z_{BS}$ should be chosen larger than most of the noise [10]. We input stochastic Gaussian noise of standard deviation $s=0.000\,25$ and $s=0.1$, then calculated the base-scale entropy of the logistic map. The results are shown in Fig. 2(d). From this we can see that the entropy also can capture all the bifurcation points except that the noise causes a small increase of entropy value in the periodic region ($3.4<r<3.57$). In the chaotic region ($3.57<r<4$) the entropy remains the same for $s<0.1$. So a small noise does not change the complexity of a chaotic signal using this method.

To further quantify the calculated results, we compared the entropy value of the logistic map with the one of stochastic white noise. For periodic, chaotic, and stochastic series, the base-scale entropy is markedly different. It is interesting that the complexities of these series are clearly different. Stochastic white noise has the largest entropy value and complexity. Contrarily the periodic series have the smallest ones. The Lorenz chaotic system was also investigated [19,20]. It also showed that the base-scale entropy captures the dynamic changes and periodic behavior very well.

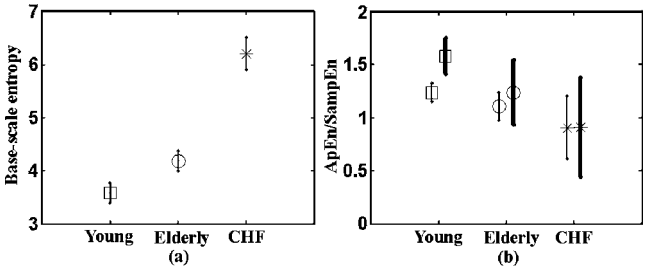

FIG. 3. (a) The comparative results of base-scale entropy for three groups. (b) The comparative results of ApEn and SampEn (thick line) for three groups. (a), (b) Values are given as mean ± standard deviation. Young (□), elderly (○), and CHF (*).

IV. APPLIED BASE-SCALE ENTROPY TO HRV SIGNALS

Heart-rate variability is the alteration in consecutive heartbeat intervals. HRV series hide a great deal of dynamic information that is related to potential mechanisms of a physiological system. Detecting and analyzing the hidden dynamical information may improve diagnosis and treatment and provide convenience for clinical applications.

We used $m=4$, $N=500$, $a=0.1$ to calculate the base-scale entropy; then we tested several values of a from 0.1 to 0.4, and the results had no significant difference. $m=3,4,5$ also had little influence on the results. Using $m=3$, even 200 data points are sufficient. We calculated the base-scale entropy for five healthy young subjects (age 27.80±4.87), five healthy old subjects (age 76.00±3.32), and 21 subjects with congestive heart failure (CHF) (age 61.14±6.92). All data sets are derived from the Fantabase Database and the Congestive Heart Failure RR Interval Database in the Physionet Database [21]. In Fig. 3(a) the result shows that the three groups can be classified completely by their base-scale entropies, through the t test, $p<0.05$. The base-scale entropy of the healthy young group denotes the optimal physiological state. The base-scale entropy of the healthy old group represents a slight deviation from the optimal youthful state due to the decoupling of components in the integrative control system with aging. Severe damage to the control system is represented by the CHF group. These individuals have profound abnormalities in cardiac function associated with pathological alterations in both the sympathetic and the parasympathetic control mechanisms that regulate beat-to-beat variability, so the base-scale entropy of the CHF group represents a big deviation from the optimal youthful state.

The harmonious physiological mechanisms have some degree of change in the aging and pathological states, which make their HRV series have a higher complexity. From short-term HRV signals, the base-scale entropy method sensitively identifies patterns generated from healthy and pathological states, as well as aging. In total, this method is feasible and effective for HRV data classification.

To further investigate the detailed information, we counted the "forbidden form" in the distribution of 4^m different forms π for three groups. The results for healthy young subjects are 228.2±6.8, healthy old subjects are 213.6±3.9, and congestive heart failure subjects are 147.5±12.1. The three groups also can be classified completely by their for-

bidden form. A small number of forbidden form denotes a rather uncertain occurrence of the m-word form. For example, the m-word forms of higher probability distribution are given as (1200), (1332), (3120), and (3312) for the healthy subjects and the probability distribution of these forms is at least 55% in all. For different subjects in the CHF group the m-word forms of higher probability distribution are different.

Figure 3(b) shows the calculated results of ApEn and SampEn for three groups. ApEn describes the rate of producing new information [10]. We use $m=2$, $r=0.2\sigma$, $N=500$ to calculate ApEn and SampEn. The three groups cannot be identified by their calculated result of ApEn and SampEn, through the t test, $p>0.05$. The causes may be the following. (i) To calculate ApEn and SampEn, one has to fix the value of a parameter that depends on the time series σ. Therefore, the results may be significantly affected by nonstationarity of the physiological time series. (ii) The series length 500 is too small to obtain a reliable convergence value. (iii) The dimension $m=2$ is too small to sensitively describe the difference of different physiological or pathological states.

V. CONCLUSION

The data sets obtained from most clinical and physiological studies usually are nonstationary, relatively short, and noisy. In our paper, we introduce the base-scale entropy as a complexity measure of time series to analyze short-term HRV series. For HRV series of 500 data points, the base-scale entropy sensitively shows the complexity dissimilarity among different physiological and pathological states, which is convenient for clinical applications.

The advantages of our method are simplicity and extremely fast calculation for very short data sets. This method enables us to analyze very short, nonstationary, and noisy data series, so the base-scale entropy can be directly applied to real-world time series.

ACKNOWLEDGMENT

We acknowledge the support of the National Natural Science Foundation of China (Grant No. 60501003).

[1] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).
[2] M. Ding, C. Grebogi, E. Ott, T. Sauer, and J. A. Yorke, Phys. Rev. Lett. **70**, 3872 (1993).
[3] A. Wolf *et al.*, Physica D **16**, 285 (1985).
[4] J. P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, Phys. Rev. A **34**, 4971 (1986).
[5] Z. Z. Wang *et al.*, Chin. Sci. Bull. **47**, 1845 (2002).
[6] J.-S. Kim *et al.*, Phys. Med. Biol. **45**, 3403 (2000).
[7] N. Perry and P.-M. Binder, Phys. Rev. E **60**, 459 (1999).
[8] P. E. Rapp, C. J. Cellucci, K. E. Korslund, T. A. A. Watanabe, and M. A. Jimenez-Montano, Phys. Rev. E **64**, 016209 (2001).
[9] J. S. Richman and J. R. Moorman, Am. J. Physiol. Heart Circ. Physiol. **278**, H2039 (2000).
[10] S. M. Pincus, Proc. Natl. Acad. Sci. U.S.A. **88**, 2297 (1991).
[11] A. Bezerianos, S. Tong, and N. Thakor, Ann. Biomed. Eng. **31**, 221 (2003).
[12] J. Kurths, A. Voss, A. Witt, P. Saparin, H. J. Kleiner, and N. Wessel, Chaos **5**, 88 (1995).
[13] N. Wessel, C. Ziehmann, J. Kurths, U. Meyerfeldt, A. Schirdewan, and A. Voss, Phys. Rev. E **61**, 733 (2000).
[14] C. S. Daw, C. E. A. Finney, and E. R. Tracy, Rev. Sci. Instrum. **74**, 915 (2003).
[15] R. Steuer *et al.*, Eur. Phys. J. B **19**, 265 (2001).
[16] W. Ebeling and R. Steuer, Stochastics Dyn. **1**, 45 (2001).
[17] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[18] C. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002).
[19] J. B. Gao and H. Q. Cai, Phys. Lett. A **270**, 75 (2000).
[20] Y. Cao, W. W. Tung, J. B. Gao, V. A. Protopopescu, and L. M. Hively, Phys. Rev. E **70**, 046217 (2004).
[21] http://www.physionet.org/physiobank/database