Phase Statistics Approach to Physiological and Financial Time Series

Ming-Chya Wu*

Ming-Chya Wu

* Research Center for Adaptive Data Analysis
National Central University
Chungli 32001, Taiwan
E-mail: mcwu@ncu.edu.tw

Dynamical systems can usually be recorded by successive recording processes, and the characteristic behaviors are caught in corresponding time series. In this article, we review the phase statistics approach introduced recently to study physiological and financial time series. The approach consists of an application of the Hilbert-Huang transform to decompose an empirical time series into a number of intrinsic mode functions (IMFs), calculation of the instantaneous phase of the resultant IMFs, and the statistics of the instantaneous phase for each IMF. We consider cardiorespiratory synchronization and phase distribution and phase correlation of stock time series as examples. The applications to other time series are also briefly discussed.

1. INTRODUCTION

Dynamical systems can usually be recorded by successive recording processes, and the characteristic behaviors are caught in corresponding time series. Among others, most of the investigations of empirical time series focus on phenomenological interpretations, such that data processing methods play a crucial role in the obtained results. However, empirical data are usually noisy, nonlinear, and nonstationary. An essential task for such studies is to process signals and pickup essential component(s) from experimental data. Even though proper filters can be used to filter out noises from real data, the capabilities and the effectiveness of the filtration are usually questionable. This is due to the fact that most of these approaches require the original time series to be stationary and/or linear. Furthermore, there is also no strict criterion to judge what is the inherent dynamics and what is the contribution of the external factors and noise in experimental data. Improper approaches might then lead to misleading results.

In this article, we review the phase statistics approach introduced recently to study physiological [1] and financial time series [2, 3, 4]. The approach mainly consists of an application of the Hilbert-Huang transform (HHT) [5] to decompose an empirical time series into a number of intrinsic mode functions (IMFs), calculation of instantaneous phase of the resultant IMFs, and the statistics of the instantaneous phase for each IMF. The HHT was introduced by Huang *et al*. [5] for the analysis of nonlinear and nonstationary time series. It consists of the empirical mode decomposition (EMD) and Hilbert spectral analysis. Unlike conventional filters, the EMD method provides an effective way to extract physical rhythms from experimental data. This advantage provides reliable interpretations from empirical time series. To illustrate the approach, here we consider cardiorespiratory synchronization (CS) and phase distribution and the phase correlation of stock time series as examples.

2. THE HIBERT-HUANG TRANSFORM

The EMD of HHT is developed around the assumption that any time series consists of simple intrinsic modes of oscillations, and the essence of the method is to identify the intrinsic oscillatory modes by their characteristic time scales in the data empirically and then decompose the data accordingly [5]. This is achieved by sifting data to generate IMFs. The IMFs introduced by the EMD are a set of well-behaved intrinsic modes and are symmetric with respect to the local mean and have the same numbers of zero crossings and extremes. Therefore, all IMFs enjoy good Hilbert transform. The algorithm to create IMFs in the EMD are rather simple, and has two main steps [1, 5]:

Step-1: Identify local extrema in the experimental data $\{x(t)\}$. All the local maxima are connected by a cubic spline line $U(t)$, which forms the upper envelope of the data. Repeat the same procedure for the local minima to produce the lower envelope $L(t)$. Both envelopes will cover all the data between them. The mean of upper envelope and lower envelope $m_1(t)$ is given by:

$$m_1(t) = \frac{U(t) + L(t)}{2} . \qquad (1)$$

Subtracting the running mean $m_1(t)$ from the original time series $x(t)$, we get the first component $h_1(t)$,

$$h_1(t) = x(t) + m_1(t). \qquad (2)$$

The resulting component $h_1(t)$ is an IMF if it is symmetric and has all maxima positive and all minima negative. An additional condition of intermittence can be imposed here to sift out waveforms with certain range of intermittence for physical consideration. If $h_1(t)$ is not an IMF, the sifting process has to be repeated as many times as it is required to reduce the extracted signal to an IMF. In the subsequent sifting process steps, $h_1(t)$ is treated as the data to repeat steps mentioned above,

$$h_{11}(t) = h_1(t) + m_{11}(t). \qquad (3)$$

Again, if the function $h_{11}(t)$ does not yet satisfy criteria for IMF, the sifting process continues up to k times until some acceptable tolerance is reached:

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{11}(t). \qquad (4)$$

Step-2: If the resulting time series is an IMF, it is designated as $c_1 = h_{1k}(t)$. The first IMF is then subtracted from the original data, and the difference r_1 given by

$$r_1(t) = x(t) - c_1(t) \qquad (5)$$

is the residue. The residue $r_1(t)$ is taken as if it were the original data, and we apply to it again the sifting process of *Step-1*.

Following the procedures of *Step-1* and *Step-2*, we continue the process to find more intrinsic modes c_i until the last one. The final residue will be a constant or a monotonic function which represents the general trend of the time series. Finally, we obtain

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n , \qquad (6)$$

$$r_{i-1}(t) - c_i(t) = r_i(t) . \qquad (7)$$

The instantaneous phase of IMF can be calculated by applying the Hilbert transform to each IMF, say the rth component $c_r(t)$. The procedures of the Hilbert transform consist of calculation of the conjugate pair of $c_r(t)$, *i.e.*,

$$y_r(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{c_r(t')}{t - t'} \, dt' , \qquad (8)$$

where P indicates the Cauchy principal value. With this definition, two functions $c_r(t)$ and $y_r(t)$ forming a complex conjugate pair, define an analytic signal $z_r(t)$:

$$z_r(t) = c_r(t) + iy_r(t) \equiv A_r(t) e^{i\phi_r(t)} , \qquad (9)$$

with amplitude $A_r(t)$ and the instantaneous phase $\phi_r(t)$ defined by

$$A_r(t) = [c_r^2(t) + y_r^2(t)]^{1/2} , \qquad (10)$$

$$\phi_r(t) = \tan^{-1}\left(\frac{y_r(t)}{c_r(t)}\right). \qquad (11)$$

Then, we can calculate the instantaneous phase according to Eqs. (8) and (11).

3. CARDIORESPIRATORY SYNCHRONIZATION

First, we present an application of the HHT to the study of CS [1]. CS is a phenomenon originating from the interactions between the cardiac and the respiratory subsystems. These interactions can lead to a perfect locking of their phases whereas their amplitudes remain chaotic and non-correlated [6]. The nature of the interactions has been extensively studied from measured data in recent years [7-15]. Recently, Schäfer *et al*. [16] and Rosenblum *et al*. [17] found that there were sufficiently long periods of hidden synchronization and concluded that the CS and respiratory sinus arrhythmia (RSA) are two competing factors in cardiorespiratory interactions. Since then, CS has been reported in young healthy athletes [16], healthy adults [18-20], heart transplant patients [18, 21], infants [22], and anesthetized rats [23]. The essential part of such investigations is the extraction of respiratory rhythms from noisy respiratory signals. A technical problem in the analysis of the respiratory signal then arises: insufficiently filtered signals may still have too many noises, and over-filtered signal may be too regular to lose the characteristics of respiratory rhythms. To overcome these difficulties, Wu and Hu [1] proposed using the HHT for such studies and got significantly reasonable results. In the implement of EMD, Wu and Hu extracted respiratory rhythms from empirical data by using the number of respiratory cycles per minute for human beings as a criterion in the sifting process of EMD [1]. They consider empirical data consisting of 20 data sets collected by the Harvard medical school in 1994 [24]. Each data set included electrocardiographic (ECG) data and respiration signals. The continuous ECG and respiration data were digitized at 250 Hz (respiratory signals were latter preprocessed to be at 5 Hz). Each heartbeat was annotated using an automated arrhythmia detection algorithm, and each beat annotation was verified by visual inspection. Each group of subjects included equal numbers of men and women.

The procedures for the analysis are as follows [1]: (i) Apply the EMD to decompose the recorded data into several IMFs. Since the respiratory signal was preprocessed to a sampling rate of 5 Hz, there should be (10-30) data points in one respiratory cycle. Thus, for example, one can use c_1: (3-6), c_2: (6-12), c_3: (12-24), ... etc. After the sifting processes of the EMD, the original respiratory data are decomposed into n IMFs c_1, c_2, ... , c_n, and a residue r_n. (ii) Visually inspect the resulting IMFs. If the amplitude of a certain mode is dominant and the waveform

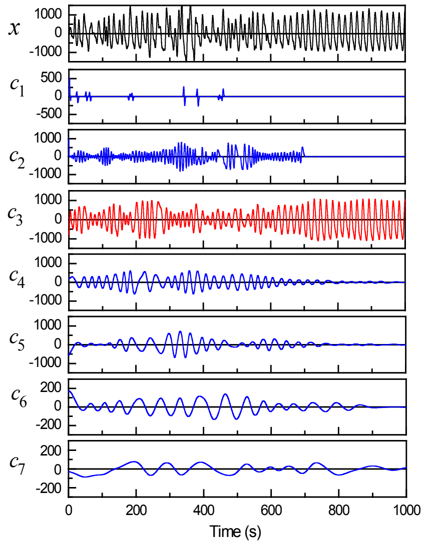

Fig. 1: Example of EMD for a typical respiratory time series. The criterion for intermittence in the sifting process is (3-6) data points per cycle for c_1. Signal x(t) is decomposed into 14 components including 13 IMFs and 1 residue; here, only the first 7 components are shown. After Ref. [1].

is well distributed, then the data are said to be well decomposed, and the decomposition is successfully completed. Otherwise, the decomposition may be inappropriate, and one has to repeat step (i) with different parameters. Fig. 1 shows a typical EMD for a respiratory time series.

After picking up one IMF as the respiratory rhythm, one can proceed in the calculation of the instantaneous phase and incorporating with heartbeat signals to construct cardiorespiratory synchrogram (CRS), which is a visual tool for inspecting synchronization. More explicitly, let us denote the phase of the respiratory signal calculated by using Eq. (11) as ϕ_r and the heartbeat as ϕ_c. If the phases of respiratory signal ϕ_r and heartbeat ϕ_c are coupled in a fashion that a cardiovascular system completes n heartbeats in m respiratory cycles, then a roughly fixed relation can be proposed. In general, there is a phase and frequency locking condition [6, 17, 18]

$$|m\phi_r - n\phi_c| \leq \text{const.} \tag{12}$$

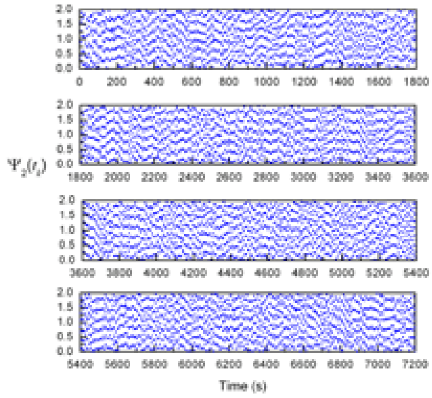

Fig. 2: CRS for a typical subject. There are about 800s synchronization at 2800-3600s, and several spells of 50-300s at other time intervals. After Ref. [1].

with m, n integer. According to Eq. (12), for the case that ECG completes n cycles while the respiration completes m cycles, it is said to be synchronization of n cardiac cycles with m respiratory cycles. Using the heartbeat event time t_k as the time frame, Eq. (12) implies the relation

$$\phi_r(t_{k+m}) - \phi_r(t_k) = 2\pi m. \tag{13}$$

Furthermore, by defining

$$\Psi_m(t_k) = \frac{1}{2\pi}\left[\phi_r(t_k) \bmod 2\pi m\right] \tag{14}$$

and plotting $\Psi_m(t_k)$ versus t_k, synchronization will result in n horizontal lines in case of n:m synchronization. By choosing n adequately, a CRS can be developed for detecting CS [16].

Example of 3:1 synchronization with n=6 and m=2 is shown in Fig. 2, where phase locking appears in several epochs, e.g. at 2800-3600s, and there is also frequency locking, e.g. at 400s, near which there are n parallel lines with the same positive slope. Here we note that if we use other filters to the same empirical data, we will have different results depending on the strength of synchronization. Wu and Hu [1] found that from the aspect of data processing that could preserve the essential features of original empirical data, the EMD approach is better than Fourier-based filtering.

The same procedures and analysis can be applied to investigate the correlation and regularity of respiratory and cardiac signals. Due to the limited space, we will not discuss this further, but refer the reader to Ref. [1] for details and other demonstrations.

4. STOCK TIME SERIES

As a second example, we present the application of the phase statistics approach to the study of stock time series [2]. The analyses are based on the DJIA and the NASDAQ from the Trade and Quotation (TAQ) database [25]. The intraday 10-minute scale values for both the DJIA and the NASDAQ spanning from August 1, 1997, through December 31, 2003, cover the whole six-and-half hours trading from 9:30 to 15:50 EST. Fig. 3 shows the DJIA and the NASDAQ index data from 1 Au-

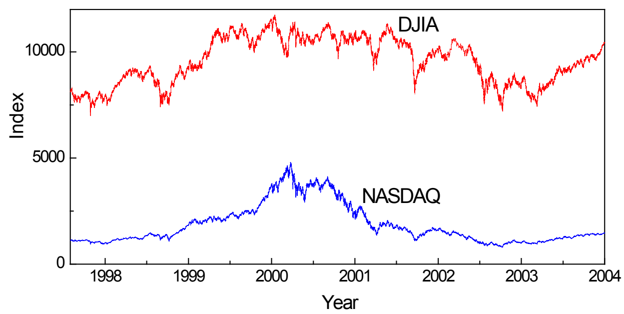

Fig. 3: DJIA and NASDAQ index data from 1 August 1997 to 31 December 2003.

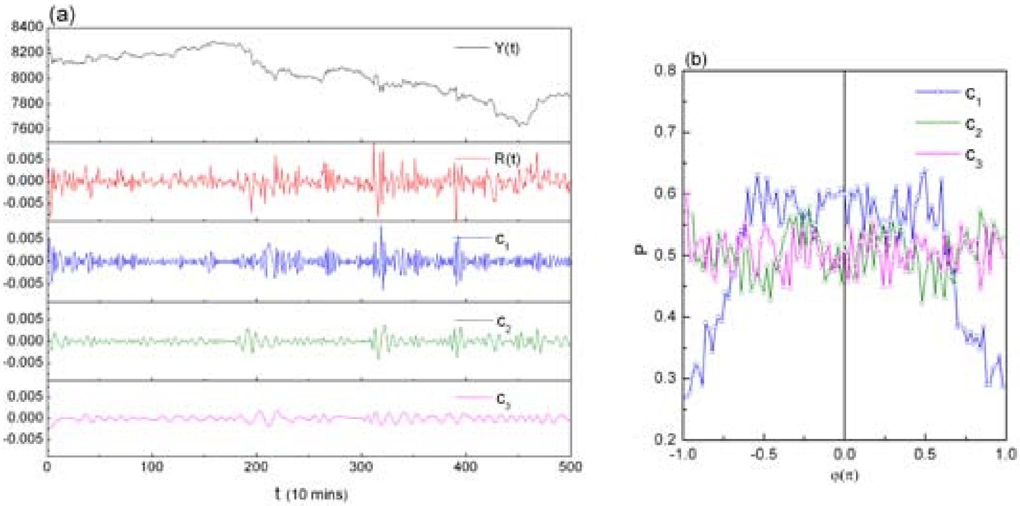

Fig. 4: (a) Intraday DJIA index and the corresponding return sampled by 10 min and the first 3 IMFs; (b) probability distribution of phases. After Ref. [2].

gust 1997 to 31 December 2003.

In the study of financial time series, one generally uses the logarithmic return. For the index values denoted by a time series $Y(t)$, the time series of logarithmic returns of an asset priced at $Y(t)$ over a time scale τ is defined as

$$R_\tau(t) = \ln Y(t) - \ln Y(t-\tau)\ , \qquad (15)$$

where τ is a multiple of the primary time sampling unit Δt (=10 min). Since the time scale τ (in unit of Δt) is a parameter used to sample time series of returns, we can take different τ for $R_\tau(t)$ to explore behaviors of the returns with intraday frequency. Because there are 39 sampling data in each trading day, we take τ = 10, 30, 130, 390 min to sample the time series of $R_\tau(t)$ for the intraday data.

The application of the phase statistics approach to investigate the return time series of the DJIA and the NASDAQ is based on the concept that the instantaneous phase can catch the characteristic features of a financial time series [2, 3, 4]. We first take the intraday returns $R_\tau(t)$ with a time sampling interval of 10 min as the primary time series and then perform the EMD to decompose $R_\tau(t)$ into 14 IMFs. The results are shown in Fig. 4(a). It is obvious that c_1 catches the main structures of $R_\tau(t)$ because the time series of $R_\tau(t)$ is mainly characterized by its highest frequency component. Similarly, we can perform the EMD on the time series with time sampling intervals of 30, 130, and 390 min. We define the probability density function (PDF) P as the normalized distribution of a measure p, which satisfies

$$\int_{-\infty}^{+\infty} P(p)\, dp = 1\ , \qquad (16)$$

where the measure p can be R_τ, phase or phase difference defined in the later discussions. We find that except for the first-IMFs of these time series, the phases of the other IMFs are randomly distributed and have equal probabilities for all possible phases, i.e., $-\pi \le \phi_1 \le \pi$, as shown in Fig. 4(b). Figs. 5(a) and 5(b) show the amplitude and the phase distributions of the first-IMFs of these time series, respectively.

The PDFs of the amplitudes for the first-IMFs are general Boltzmann distributions. Among these, the phase distribution is quite interesting. Most phases of the first-IMFs locate at $-0.5\pi \le \phi_1 \le 0.5\pi$. The clustered distribution of phase originates from abrupt changes in the behaviors of the index time series, which is a nature of a time series with intermittency close to the sample

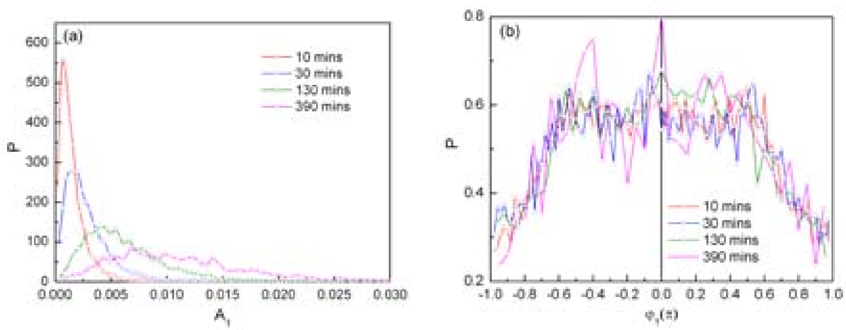

Fig. 5: Probability distributions of (a) amplitudes and (b) phases for the first-IMFs of the returns of the DJIA index sampled by 10, 30, 130, and 390 min. After Ref. [2].

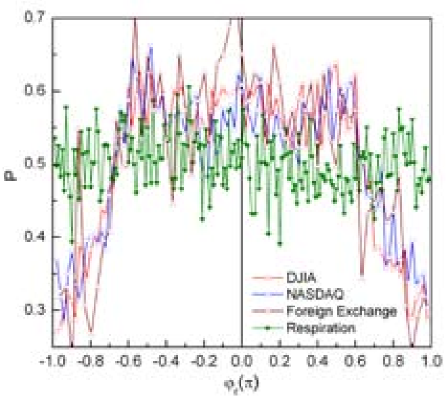

Fig. 6: Probability distributions of phases for the first-IMFs of the returns for the DJIA and the NASDAQ indices sampled by 10 min, and the third IMF of a typical respiratory time series. After Ref. [2].

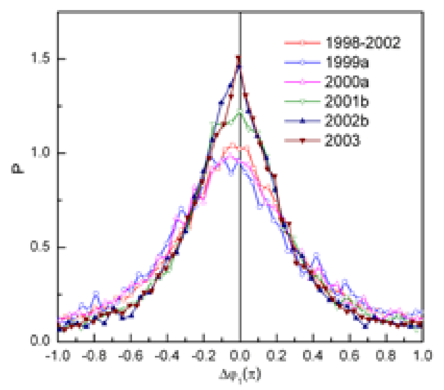

Fig. 7: Probability distributions of phase differences between the first-IMFs of returns of the DJIA and the NASDAQ indices for 1998-2002 and for certain periods and events. After Ref. [2].

time scale τ. These behaviors exist for all sample time scales of the intraday data. It is remarkable that the distributions of phases are the same in spite of the compositions of their stocks. This implies that it is a characteristic behavior of such kinds of time series. For comparison, we calculate the phase distribution of a typical respiratory time series and compare the PDFs of phases for returns of the DJIA and the NASDAQ indices and for the respiratory time series in Fig. 6. From the results, it is apparent that the return time series and the respiratory time series belong to different classes.

To investigate the correlative behaviors between two stocks, we apply the HHT to calculate instantaneous phases of several epochs of the return time series of DJIA and NASDAQ indices. Here we further define phase differences of the first-IMFs for different indices. Taking DJIA as a reference, we define phase difference $\Delta\phi_1$ as

$$\Delta\phi_1 = \phi_1(\text{NASDQA}) - \phi_1(\text{DJIA}), \qquad (17)$$

and then calculate PDFs for certain epoches and events. Fig. 7 shows the PDFs of phase differences between the first-IMFs of returns of the two indices for 1998-2002, the first half year of 1999 (indicated by 1999a), the first half year of 2000 (indicated by 2000a), the last half years of 2001 and 2002 (indicated by 2001b and 2002b, respectively), and the whole year of 2003. We find that there is a remarkable change in the behavior of the trading activities both in the DJIA and the NASDAQ since 9/11 attack. More specifically, Fig. 7 shows that there were more correlative activities after 9/11 so that the distribution functions of 2001b and 2002b were quite different from those before 9/11. There was a similar spectrum in the year 2003, implying the scenario persisted in later trading activities. This was seen as faster information transmission and stronger event dependence in the stock markets after 9/11 [2].

5. DISCUSSIONS

We have briefly explored the scheme of the phase statistics approach and its applications to the study CS and the analysis of stock time series. The remarkable advantage of the HHT used in this approach is that it can catch primary structures of intrinsic rhythms from empirical data based on its adaptive feature [5, 26]. This property is especially suitable for performing phase statistics on empirical time series. By imposing intermittency criteria based on physical conditions revealed by empirical time series, this feature also allows us to effectively keep the signal structures. Furthermore, the introduction of IMFs in the EMD provides a reasonable definition of the instantaneous phase which is also helpful for the implement of phase statistics approach.

Recently, the phase approach has been successfully applied to the study of foreign exchange time series [3] and the identification of fatal ventricular fibrillation (VF) from human ECG data [27]. Wu [3] found that the correlation between the USD/DEM and USD/JPY exchange rates was stronger during 1986-1989 than 1990-1993, which was consistent with the observations from cross-correlation calculation. The scenario was explained by event of unification of East Germany and West Germany and the era of bubble economy of Japan in early 1990s [3]. Moreover, Wu *et al.* [27] applied the phase statistics approach to the investigations of human VF time series. By specifying phase distribution patterns for fatal and non-fatal VFs, they found fatal VF can be identified with a probability higher than 80% [27]. According to the impressive achievements of the application of the phase statistics approach to time series analysis with the aid of the HHT, we expect that the approach presented herein also can be useful for statistical analysis of other time series, such as time series of temperature variation, seismic time series, biological systems [28], and other social models.

This work was partially supported by the National Science Council of the Republic of China (Taiwan) under Grant No. NSC 95-2119-M-002-001. The author thanks collaborators for their collaborations on the works of Refs. [1, 2] and [27, 28].

REFERENCES

[1] M.-C. Wu and C.-K. Hu, *Phys. Rev. E* **73**, 051917 (2006).
[2] M.-C. Wu, M.-C. Huang, H.-C. Yu, and T.-C. Chiang, *Phys. Rev. E* **73**, 016118 (2006).
[3] M.-C. Wu, *Physica A* **375**, 633 (2007).
[4] M.-C. Wu, *J. Korean Phys. Soc.* **50**,

304 (2007).
[5] N.-E. Huang, Z. Shen, S.-R. Long, M.-C. Wu, H.-H. Shih, Q. Zheng, N.-C. Yen, C.-C. Tung, and H.-H. Liu, *Proc. R. Soc. Lond. A* **454**, 903 (1998).
[6] P. Tass, M. G. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, A. Schnitzler, and H. J. Freund, *Phys. Rev. Lett.* **31**, 3291 (1998).
[7] A. C. Guyton, *Text book of medical physiology* (Saunders, Philadelphia, PA, 1991), 8th ed.
[8] L. Bernardi, F. Salvucci, R. Suardi, P. L. Solda, A. Calciati, S. Perlini, C. Falcone, and L. Ricciardi, *Cardiovasc. Res.* **24**, 969 (1990).
[9] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, *Fluctuation and Noise Lett.* **4**, L53 (2004); M. G. Rosenblum and A. S. Pikovsky, *Phys. Rev. Lett.* **92**, 114102 (2004).
[10] T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).
[11] M. Paluš and A. Stefanovska, *Phys. Rev. E* **67**, 055201(R) (2003).
[12] J. Jamsek, A. Stefanovska and P. V. E. McClintock, *Phys. Med. Bio.* **49**, 4407 (2004).
[13] M. Richter, T. Schreiber, and D. T. Kaplan, *IEEE Eng. Med. Biol. Mag.* **45**, 133 (1998).
[14] R. Hegger, H. Kantz, and T. Schreiber, *Chaos* **9**, 413 (1999).
[15] H. Kantz and T. Schreiber, *IEE Proceedings – Science Measurement and Technology* **145**, 279 (1998).
[16] C. Schäfer, M. G. Rosenblum, J. Kurths, and H. H. Abel, *Nature* (London) **392**, 239 (1998); C. Schäfer, M. G. Rosenblum, H. H. Abel, and J. Kurths, *Phys. Rev. E* **60**, 857 (1999).
[17] M. G. Rosenblum, J. Kurths, A. Pikovsky, C. Schäfer, P. Tass, and H. H. Abel, *IEEE Eng. Med. Biol. Mag.* **17**, 46 (1998).
[18] E. Toledo, M. G. Roseblum, C. Schäfer, J. Kurhts, and S. Akselrod, "Quantification of cardiorespiratory synchronization in normal and heart transplant subjects", in *Proc. of Int. Symposium on Nonlinear Theory and its Applications* (Presses polytechniques et universitaries romandes, Lausanne, 1998), Vol. 1, pp. 171-174.
[19] E. Toledo, M. G. Rosenblum, J. Kurths, and S. Akselrod, "Cardiorespiratory synchronization: is it a real phenomenon?", in *Computers in Cardiology* (IEEE Computer Society, Los Alamitos, CA, 1999),Vol. 26, pp. 237-240.
[20] M. B. Lotric and A. Stefanovska, *Physica A* **283**, 451 (2000).
[21] E. Toledo, S. Akselrod, I. Pinhas, and D. Aravot, *Med. Eng. Phys.* **24**, 45 (2002).
[22] R. Mrowka and A. Patzak, *Int. J. of Bifurcation and Chaos* **10**, 2479 (2000).
[23] A. Stefanovska, H. Haken, P. V. E. McClintock, M. Hozic, F. Bajrovic, and S. Ribaric, *Phys. Rev. Lett.* **85**, 4831 (2000).
[24] N. Iyengar, C.-K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, *Am. J. Physiol.* **271**, 1078 (1996). Data are available from http://physionet.org/physiobank/database/ fantasia/.
[25] We use the TAQ database published by the New York Stock Exchange to construct the intraday returns.
[26] The power and confidence limit of the EMD method for various empirical time series have been justified and discussed thoroughly in Ref. [5] and N.-E. Huang, M.-C. Wu, S.-R. Long, S.-S. P. Shen, W. Qu, P. Gloersen, and K.-L. Fan, *Proc. R. Soc. Lond. A* **459**, 2317 (2003).
[27] M.-C. Wu, E. Watanabe, Z. R. Struzik, C.-K. Hu, and Y. Yamamoto, "Identification of fatal ventricular fibrillation patterns from phase statistics of human ECG data", manuscript submitted (2007).
[28] M.-C. Wu, C.-K. Hu, H.-M. Chen, and T.-Y. Tsong, "Local hydrophobicity in protein folding: Theoretical model and experimental test with staphylococcal nuclease and its mutants", manuscript submitted (2007).