Information transfer at multiple scales

Max Lungarella* and Alex Pitti
ERATO Asada Synergistic Intelligence Project, JST, The University of Tokyo, 113-8656 Tokyo, Japan

Yasuo Kuniyoshi
Department of Mechano-Informatics, The University of Tokyo, 113-8656 Tokyo, Japan
(Received 20 April 2007; published 27 November 2007)

In the study of complex systems a fundamental issue is the mapping of the networks of interaction between constituent subsystems of a complex system or between multiple complex systems. Such networks define the web of dependencies and patterns of continuous and dynamic coupling between the system's elements characterized by directed flow of information spanning multiple spatial and temporal scales. Here, we propose a wavelet-based extension of transfer entropy to measure directional transfer of information between coupled systems at multiple time scales and demonstrate its effectiveness by studying (a) three artificial maps, (b) physiological recordings, and (c) the time series recorded from a chaos-controlled simulated robot. Limitations and potential extensions of the proposed method are discussed.

DOI: 10.1103/PhysRevE.76.056117 PACS number(s): 89.75.−k, 05.45.Tp, 89.70.+c

I. INTRODUCTION

The question often occurs of how to identify and characterize information structure and hidden dependencies between the components of a complex dynamical system or between multiple complex systems, given only a set of simultaneously recorded (typically multivariate) time series. In particular, it may be required to identify causal dependencies or quantify flow of information. Such flow can be unidirectional—i.e., time series X_t affects time series Y_t, while Y_t has no influence on X_t— or (in the general case) bidirectional. Moreover, because complex systems operate across multiple spatial and temporal scales, it is likely that information flow will exhibit a multiscale structure.

Cross-correlation is a common way of evaluating the statistical dependencies among data sets; it is not only efficiently computed but has also an appealing and natural interpretation. Cross-correlation, however, is a second-order statistics handling merely linear dependencies. Mutual information provides an attractive way of circumventing the restrictions of cross-correlation because it is sensitive to higher-order relationships, both linear and nonlinear ones. Yet, mutual information is a symmetric measure and fails to detect directional (asymmetric) flow of information unless one of the time series is delayed [1]. Measures of interdependence aimed at overcoming the limitations of mutual information have recently been introduced, e.g., extended Granger causality [2,3], nonlinear Granger causality [4], similarity index [5], predictability improvement [6], and transfer entropy [7]. All these measures, however, do not account for features related to structure on scales other than the shortest one.

Here, we introduce a multiscale extension of transfer entropy [7] as a tool for understanding the internal dynamics of complex systems. By projecting a time series into the wavelet space, a new set of variables is obtained, whose statistics allow us to extract causal dependencies between the variables. In what follows, we show that our wavelet-based approach augments "sample-based" embedding vector approaches by disclosing previously undetectable causal structure in potentially nonstationary and discontinuous signals. Our analysis is applicable to a wide range of nonstationary physical and biological signals, regardless of whether the underlying fluctuations have stochastic origins or arise from nonlinear dynamical processes.

The paper is organized as follows. We first introduce a wavelet-based multiscale extension of transfer entropy. Transfer entropy builds upon conditional entropy, which in contrast to the standard form of mutual information, or delayed-related measures, allows one to distinguish actually "transported" from shared information. We then illustrate the application of our method by studying three synthetic data sets, a physiological time series, and data collected using a simulated robot. Next, we discuss the results, point to some of the shortcomings of the extension, and eventually draw some conclusions.

II. METHODS

In this section we give the mathematical background for the proposed measure. We first outline some formalism and then describe the method. The method was implemented in MATLAB R7.2 (MATHWORKS, MA) and the MATLAB wavelet toolbox using purpose-built script files.

A. Transfer entropy

We are interested in detecting directed exchange of information between two distinct dynamical systems X and Y and in comparing their mutual influence. A good approach to obtain knowledge about asymmetric dependencies between coupled systems (or the direction and strength of the coupling between subsystems) is to measure to which extent its individual components contribute to information production and at what rate they exchange information among each

*Corresponding author. lunga@ifi.uzh.ch

 ©2007 The American Physical Society

other. Transfer entropy [7] is an appropriate starting point because it is an information theoretic measure which not only shares some of the properties of mutual information but also takes the dynamics of information transport into account. It is a specific version of the mutual information for conditional probabilities, but unlike mutual information it is designed to detect information transferred between two systems, separate for both directions, and conditional to common history and input signals.

Let $X_t=\{x_t\}_{t=1,\ldots,N}$ and $Y_t=\{y_t\}_{t=1,\ldots,N}$ denote bivariate stationary time series simultaneously measured from X and Y. We start by reconstructing the dynamics of the two time series by the method of time delay embedding [8,9]. The state (or embedding) vector of dimension m of X_t is formed by delayed past scalar observations of the time series:

$$\boldsymbol{x}_t^{m,d} = (x_t, x_{t-d}, x_{t-2d}, \ldots, x_{t-(m-1)d})^T, \tag{1}$$

where d is the time delay (or lag) between successive elements of the state vector, and $t=\eta+1,\ldots,N$ with embedding window $\eta=(m-1)d$. If not otherwise stated in this paper $d=1$, and we use the following expression:

$$\boldsymbol{x}_t^{m} = (x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-m+1})^T. \tag{2}$$

Similarly, we can define $\boldsymbol{y}_t^{m,d}$ and $\boldsymbol{y}_t^{m}$. Note that the state vectors $\boldsymbol{x}_t^{m,d}$ and $\boldsymbol{y}_t^{m,d}$ are points in the m-dimensional spaces X and Y. To reconstruct the state spaces of the time series, the embedding dimension m and the time delay d have to be determined. The choice of both parameters depends on the dynamics of the underlying data.

Essentially, transfer entropy measures the deviation from the generalized Markov property

$$p(x_{t+1}|\boldsymbol{x}_t^k) = p(x_{t+1}|\boldsymbol{x}_t^k, \boldsymbol{y}_t^l), \tag{3}$$

where $p(\cdot|\cdot)$ denotes the transition probability, and k and l are the dimensions of the two delay vectors. If the deviation from the generalized Markov process is small, i.e., $p(x_{t+1}|\boldsymbol{x}_t^k) \approx p(x_{t+1}|\boldsymbol{x}_t^k, \boldsymbol{y}_t^l)$, then it can be assumed that the state vectors of space Y have no (or little) influence on the transition probabilities of the state vectors of space X. If the deviation is large, however, then the assumption of the Markov process is not valid. The incorrectness of the assumption can be quantified by the transfer entropy which is formulated as a special instance of the Kullback-Leibler entropy between $p(x_{t+1}|\boldsymbol{x}_t^k)$ and $p(x_{t+1}|\boldsymbol{x}_t^k, \boldsymbol{y}_t^l)$ [7]:

$$T_{Y\to X} = \sum_{x_{t+1},\boldsymbol{x}_t^k,\boldsymbol{y}_t^l} p(x_{t+1},\boldsymbol{x}_t^k,\boldsymbol{y}_t^l) \log \frac{p(x_{t+1}|\boldsymbol{x}_t^k,\boldsymbol{y}_t^l)}{p(x_{t+1}|\boldsymbol{x}_t^k)}, \tag{4}$$

where x_t and y_t represent the states of time series X_t and Y_t at time t. An appealing feature of the transfer entropy is that—unlike mutual information or conditional entropy—it is explicitly nonsymmetric under the exchange of X_t and Y_t (a similar expression exists for $T_{X\to Y}$). It can thus be used to detect coupling and directional transport between two systems. In other words, the transfer entropy represents the information about a future observation x_{t+1} obtained from the simultaneous observation of k past values of X_t and l past values of Y_t, after discarding the information about the future of X_t obtained from k past values of X_t alone.

The transfer entropy is obtained by approximating the joint probabilities $p(x_{t+1},\boldsymbol{x}_t^k,\boldsymbol{y}_t^l)$ through kernel density estimation and by coarse-graining the continuous state space at a resolution r. For computational reasons we set $k=l=1$ and obtain:

$$\hat{p}_r(z_t) = \frac{1}{L}\sum_{t'} \Theta(r - \|z_t - z_{t'}\|), \tag{5}$$

where $z_t=(x_{t+1},x_t,y_t)^T$, $\|\cdot\|$ is the Euclidean distance, and L is the number of pairs (t,t') (which are selected to exclude dynamically correlated pairs). As in [7], we use a step kernel: $\Theta(x>0)=1$; $\Theta(x\leq 0)=0$ (that is, the probabilities are evaluated as normalized histograms). Alternative kernels can be found in [10]. The conditional probabilities required for Eq. (4) are then calculated from the joint probabilities.

When the available data is limited (number of samples $N<1000$) and the coupling between the time series is small, transfer entropy suffers from a finite sample effect (in particular for small resolutions, $r<0.05$) which makes the assessment of the significance of the obtained values difficult [11]. To counter this problem, a slightly modified "finite sample"-corrected variant of transfer entropy, called effective transfer entropy, has been proposed [12]. Note that we do not deal with the finite sample problem here. For all our experiments $N>1000$ and $r>0.05$, and we can safely assume that our results are affected by a finite sample problem just to a small degree.

B. Wavelet-based extension

Multiresolution wavelet analysis is a useful technique for analyzing signals at multiple scales, even in the presence of nonstationarities which often obscure such signals [13,14]. Wavelet analysis is known for its superiority to both conventional and short-time (windowed) Fourier analysis when it comes to time and frequency localization [14,15], and thus lends itself particularly well as a tool for analyzing nonstationary and discontinuous signals. Wavelet coefficients not only provide a compact representation of the power distribution in time and scale (or frequency), but they also capture information about the higher frequency oscillatory properties and long-term trends.

A doubly indexed family of wavelet coefficients $V_{s,\tau}$ can be extracted from X_t by convolving X_t with a scaled and translated version of a function called the mother wavelet (or prototype function) $\psi(\eta)$ (an analogous family $W_{s,\tau}$ can be defined for Y_t):

$$V_{s,\tau} = \frac{1}{\sqrt{s}}\sum_{t=1}^{N} X_t \psi^*\left(\frac{t-\tau}{s}\right), \tag{6}$$

where the $(^*)$ denotes the complex conjugate, $s>0$ is the scale dilation parameter (corresponding to the width of the wavelet), and $\tau\in R$ is the translation parameter indicating the location of the wavelet function as the prototype function is shifted through the signal (the point of reference of the con-

volution). The scale s determines the width of the wavelet function and hence its resolution. Smaller scales correspond to more rapid variations and therefore to higher frequencies; that is, the larger the scale, the wider is the wavelet, and the more the details are smeared out. In this regard, the wavelet transform resembles a series of bandpass filters. Each filter may be assigned an equivalent Fourier period which gives an idea of the frequency content of the signal analyzed. This is done by substituting a cosine wave of a known frequency into Eq. (6) and by computing the scale that matches the maximum wavelet power spectrum.

Many choices for the mother wavelet functions exist that differ according to symmetry, width of support, and orthogonality [13]. Here, we use the (complex) MORLET wavelet as it is characterized by a good trade-off between time and frequency localization [36]:

$$\psi(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^2/2}, \tag{7}$$

where ω_0 is the normalized frequency and is here taken to be $\omega_0 \in [5,6]$. With reference to the MATLAB wavelet toolbox, the mother wavelets used are *cmor2–0.9549* ($\omega_0=6$; synthetic and physiological data; Secs. III A and III B) and *cmor2–0.7958* ($\omega_0=5$; robot data; Sec. III C). For the MORLET wavelet the relationship between scale and equivalent Fourier period is $\lambda = 1.03$ s.

We extend transfer entropy by mapping the information transfer in the space spanned by the wavelet coefficients. The formulation is straightforward. First, the wavelet coefficients $V_{s,\tau}$ and $W_{s,\tau}$ are extracted from the time series X_t and Y_t for the scales $s=s_0,\ldots,s_n$. The mth scale is $s_m = s_0 2^{m/V}$, where $s_0 \in R^+ - \{0\}$ and V is a positive integer number (corresponding to the number of scales for each octave). The smallest resolvable scale s_0 ($m=0$) should be chosen so that the equivalent Fourier period $\lambda(s_0) \approx 2T_s$ (where T_s is the sampling period). For $V \geq 4$ the wavelets constructed are quasi-orthogonal and the scales quasi-independent [13,14] (for all our analyses, $V \geq 8$). In analogy to Eq. (2), we define two "embedding" vectors using the wavelet coefficients:

$$\boldsymbol{V}_{s,\tau} = (V_{s_1,\tau}, \ldots, V_{s_n,\tau})^T, \tag{8}$$

$$\boldsymbol{W}_{s,\tau} = (W_{s_1,\tau}, \ldots, W_{s_n,\tau})^T. \tag{9}$$

Assuming that $V_{s,\tau}$ and $W_{s,\tau}$ are "lag-stationary" (the lag is denoted by τ), one can express the multiscale extension of Eq. (4) for $k=l=1$ as

$$T_{Y\to X}(s) = \sum p(V_{s,\tau+1}, V_{s,\tau}, W_{s,\tau}) \log \frac{p(V_{s,\tau+1} | V_{s,\tau}, W_{s,\tau})}{p(V_{s,\tau+1} | V_{s,\tau})}, \tag{10}$$

$$T_{X\to Y}(s) = \sum p(W_{s,\tau+1}, W_{s,\tau}, V_{s,\tau}) \log \frac{p(W_{s,\tau+1} | W_{s,\tau}, V_{s,\tau})}{p(W_{s,\tau+1} | W_{s,\tau})}, \tag{11}$$

where $s>0$ indicates the scale. If $T_{Y\to X}(s) - T_{X\to Y}(s) > 0$, information flows from Y to X at scale s, and vice versa if $T_{X\to Y}(s) - T_{Y\to X}(s) > 0$, information flows from X to Y.

One pitfall of this measure (as well as of other measures [16]) is that it does not readily allow us to assess the significance of the result. This problem can be addressed with the concept of surrogates [17]. Here, we generate a set of surrogates of $\{\boldsymbol{V}_{s,\tau}, \boldsymbol{W}_{s,\tau}\}$ by permuting the temporal indices of $\boldsymbol{W}_{s,\tau}$ while keeping those of $\boldsymbol{V}_{s,\tau}$ fixed. This operation effectively destroys the realizationwise pairings of the times series while preserving their original statistical structure. Applying transfer entropy to the original time series and to q surrogate data sets—each obtained through a different random permutation—yields $T(s,q)$ and allows testing the null hypotheses that X_t and Y_t are extracted from decoupled systems. As in [18] or [12], however, our primary use of surrogates is not to test this null hypothesis but rather for an offset correction, which can be interpreted as a significance threshold. We finally define the offset-corrected wavelet-based transfer entropy as

$$T_c(s) = T(s) - \max_q \{T(s,q)\}. \tag{12}$$

In what follows, $T_{c,X\to Y}$ will indicate "offset-corrected" information transferred from X to Y, whereas $T_{c,Y\to X}$ will indicate information transferred in the opposite direction. In all the following numerical experiments, $q=10$.

III. EXPERIMENTS

We tested our method on five data sets: three unidirectionally coupled chaotic maps, one physiological time series, and a data set generated using a robot.

A. Synthetic data

As a first experiment, for simplicity, we studied the system formed of two unidirectionally coupled autoregressive processes of first order where one process contains a cosine:

$$x_{t+1} = 0.7x_t + 0.7\cos(0.3t) + n_t^{(x)}(0,\sigma^2), \tag{13}$$

$$y_{t+1} = 0.7y_t + n_t^{(y)}(0,\sigma^2) + ex_t, \tag{14}$$

where the parameter $e \in [0,1]$ regulates the coupling strength, and $n_t^{(x)}$ and $n_t^{(y)}$ are independent Gaussian random processes with zero mean and variance $\sigma^2=0.2$. By construction, the two time series are unidirectionally coupled with time series X_t influencing time series Y_t. For each coupling, we generated a bivariate time series (X_t, Y_t) (random initial conditions extracted from a normal distribution with zero mean and unit variance; $N=5\times10^4$ samples, the first 2×10^4 samples are discarded).

The parameters used for the analysis were $r=0.125$, $\omega_0=6$, $s_0=0.5$, $V=10$, and $n=64$ scales. Figure 1 displays the offset-corrected directed information exchange between the two maps $\Omega_c = T_{c,x_t\to y_t} - T_{c,y_t\to x_t}$ as a function of the coupling strength e and scale s. Because for $e=0$ the two processes are independent, $\Omega_c \approx 0$ for any value of s. For increasing values of e the influence of X_t on Y_t increases as demonstrated by

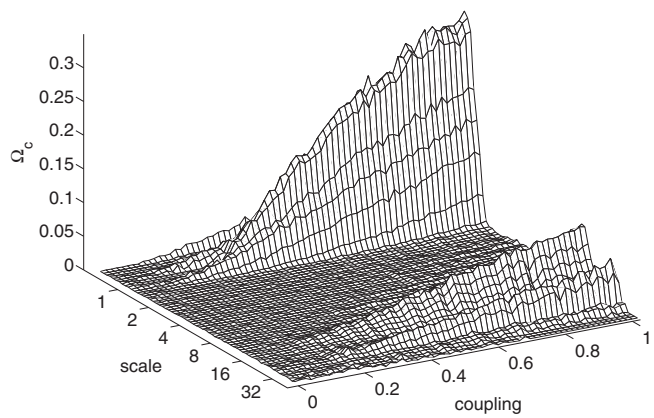

FIG. 1. Wavelet-based transfer entropy for coupled autoregressive processes (x_t^1 contains a cosine term). Displayed is $\Omega_c = T_{c,x_t^1 \to x_t^2} - T_{x_t^2 \to x_t^1}$ as a function of scale s and coupling strength e ($\omega_0=6$, $s_0=0.5$, $V=10$).

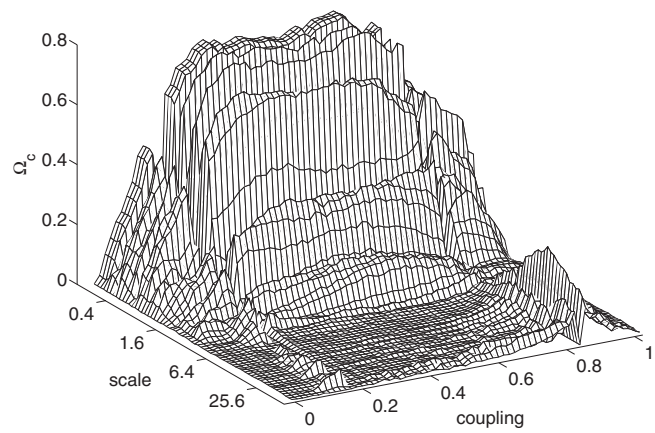

FIG. 2. Wavelet-based transfer entropy for coupled Ulam maps. Displayed is $\Omega_c = T_{x_t^1 \to x_t^2} - T_{c,x_t^2 \to x_t^1}$ as a function of scale s and coupling strength e ($\omega_0=6$, $s_0=0.2$, $V=8$).

the larger values of Ω_c. In particular, one can observe the effect of the cosine on the dynamics of the coupled system: a slowly increasing ridge, rather well-localized in scale, in the scale interval [12,28]. We also estimated the information exchange as a function of the coupling e using the original (nonmultiscale) definition of transfer entropy [Fig. 4(a); see also [11]]. The comparison between the "traditional" and the multiscale approach reveals the advantage of the latter; that is, a more fine-grained detection of information flow (or, causal dependencies) potentially allowing a better understanding of the mechanisms underlying the dynamics of a particular system.

In our second numerical experiment, we studied a one-dimensional ring lattice of 100 unidirectionally coupled Ulam maps with periodic boundary conditions:

$$x_{t+1}^l = f[ex_t^{l-1} + (1-e)x_t^l], \tag{15}$$

$$f(u) = 2 - u^2, \quad l = 1, \ldots, 100, \tag{16}$$

where the parameter $e \in [0,1]$ denotes the coupling strength, l is the lattice index, and t the time index. Note that by construction, information can be transported only in the direction of increasing l. For each coupling, we generated the two time series $X_t^1(l=1)$ and $X_t^2(l=2)$ with a sample size of $N=2\times 10^4$, after 2×10^4 steps of transients (random initial conditions).

The analysis was conducted using the following parameters: $r=0.125$, $\omega_0=6$, $s_0=0.2$, $V=8$, and $n=64$ scales. Figure 2 shows the directed exchange of information $\Omega_c = T_{c,x_t^1 \to x_t^2} - T_{c,x_t^2 \to x_t^1}$ as a function of e and s. As evident from the figure, Ω_c correctly detects the gradual increase of coupling strength for increasing values of e, reflecting the causality of the system. This result is to be expected by the way the ring lattice was constructed, and for high values of the scale index is consistent with what reported in the literature [6,7]. For $e=0.18$ and 0.82, there is a sharp drop in the flow of information (at all scales), and the estimated coupling is zero. For these two values of e, map ($l=1$) and map ($l=2$) are in a (generally) synchronized state, that is, driver and response system are indistinguishable from each other. As in the previous numerical experiment, also in this case, we calculated the amount of information exchanged between the two selected neighboring maps using the original version of transfer entropy [Fig. 4(b)] and compared it to the proposed multiscale approach. Although for coupling values $e<0.5$ and for scales $s<6.4$ the trends are qualitatively similar, one can observe a discrepancy for $e>0.5$ and $s<6.4$: at low scales the influence of a high coupling value is less pronounced. For $s>6.4$, however, the relative influence of a higher coupling value is larger, hinting at a low frequency information flow between the two maps.

In our third numerical experiment, we studied two unidirectionally coupled Henon maps:

$$x_{t+2} = 1.4 - x_{t+1}^2 + 0.3x_t, \tag{17}$$

$$y_{t+2} = 1.4 - \{ex_{t+1} + (1-e)y_{t+1}\}y_{t+1} + 0.3y_t. \tag{18}$$

As in both previous systems, $e \in [0,1]$ is the coupling parameter. In this particular system, the maps synchronize for $e>0.7$. For such values of the coupling, no quantifiable difference between X_t and Y_t exists, and the information transferred from one map to the other is zero. The wavelet-based transfer entropy was calculated using $N=2\times 10^4$ samples after discarding the 2×10^4 initial samples as transients.

The parameters for the analysis were $r=0.125$, $\omega_0=6$, $s_0=0.2$, $V=10$, and $n=64$ scales. The directed exchange of information $\Omega_c = T_{c,x_t \to y_t} - T_{c,y_t \to x_t}$ is shown in Fig. 3. The measure increases with the coupling strength e up to a critical point $e^*=0.7$. For $e>e^*$, the two maps synchronize at all scales and $\Omega_c \approx 0$, as expected. As in the case of the coupled Ulam maps, we note that the trends of both indices remain consistent for all couplings and scales, i.e., $T_{c,x_t \to y_t} > 0$ and $T_{c,y_t \to x_t} \approx 0$, reflecting the causality in the system. Our findings are consistent with the results reported in [2]. As in the two previous numerical experiments, also here we compared our multiscale approach with a "single scaled" one [Fig. 4(c)]. There is a good match between the amount of ex-

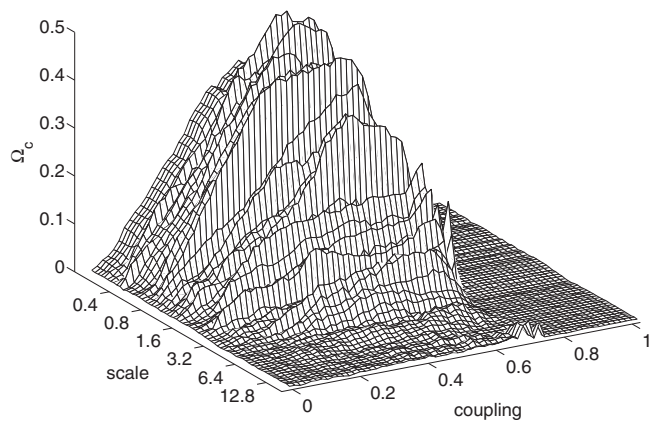

FIG. 3. Wavelet-based transfer entropy for coupled Henon maps. Displayed is $\Omega_c = T_{c,x\to y} - T_{y\to x}$ as a function of scale s and coupling strength e ($\omega_0=6$, $s_0=0.2$, $V=10$).

changed information as a function of the coupling measured by the wavelet-based multiscale extension (at any single scale s) and the original version of transfer entropy.

B. Physiological data

For our fourth numerical experiment, we used a bivariate time series extracted from a publicly available physiological data set [19] consisting of the breath rate (measured as lung pressure) and the instantaneous heart rate of a sleeping human patient suffering from a breathing disorder. Under normal, physiological conditions, the heart rate is modulated by respiration through a process known as respiratory sinus arrhythmia (RSA). Sleep apnea (a pathological condition characterized by brief interruptions of breathing during sleep) affects the normal process of RSA disturbing the usual (physiological) patterns of interaction and feedback among heart rate, respiration, and blood oxygen concentration. As a consequence, the control of heart rate by respiration becomes unclear. We conjecture that the application of wavelet-based transfer entropy may provide additional insights into the nature of the interaction. In particular because it is likely that the dynamics of mutual interaction among physiological variables act on different (*a priori* unknown) time horizons.

Figure 5 reproduces the data set that we analyzed. The figure indicates that bursts of breath and beat-to-beat alterations in heart rate [heart rate variability (HRV)] are interdependent (samples no. 305–4305; approximately 32 min). The parameters used for the analysis were $r=0.16$, $\omega_0=6$, $s_0=0.9$, $V=60$, and $n=256$ scales. Figure 6 shows that our multiscale approach finds bidirectional flow of information (from heart rate to breath rate, and vice versa). The graph also illustrates that at lower scales ($s\in]0.0,1.463]$, corresponding to the frequency interval 0.321–2.0 Hz) the flow of information from heart rate to breath rate ($T_{c,H\to B}$) is larger than the one from breath rate to heart rate ($T_{c,B\to H}$). This result is consistent with the findings reported in [2,7,20]. Similarly, at higher scales ($s\in[5.5,8.0]$; 0.058–0.086 Hz), the causal influence of heart rate on breath rate is larger than vice versa ($T_{c,H\to B} > T_{c,B\to H}$). For

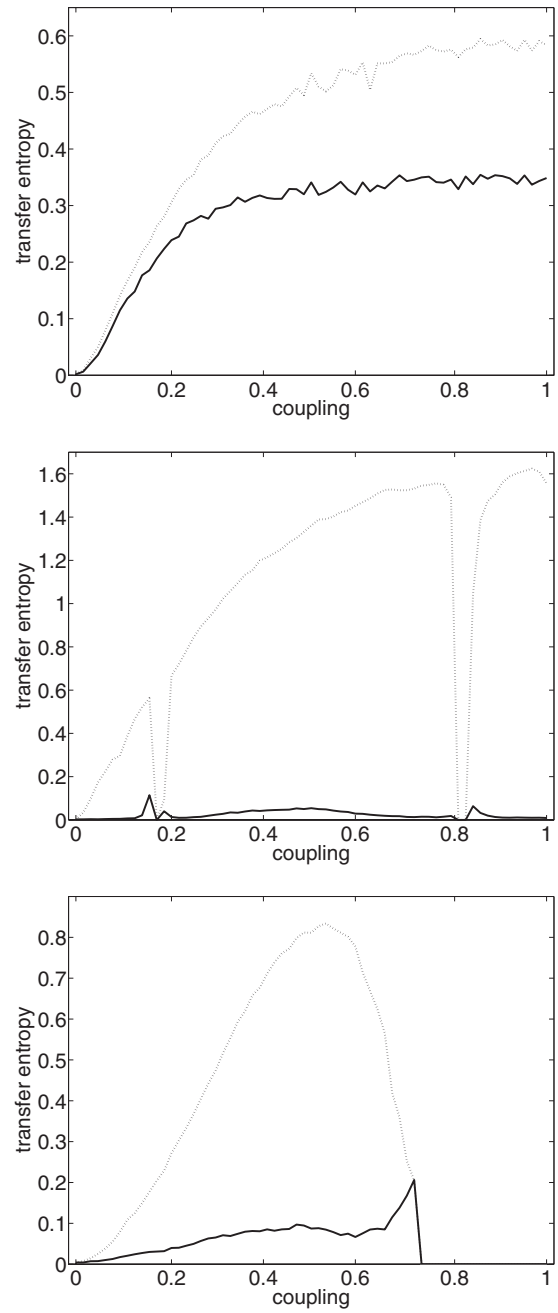

FIG. 4. Transfer entropy as a function of the coupling e for (a) coupled autoregressive processes ($T_{x_t\to y_t}$: dotted, $T_{y_t\to x_t}$: solid); (b) coupled Ulam maps ($T_{x_t^1\to x_t^2}$: dotted, $T_{x_t^2\to x_t^1}$: solid); and (c) coupled Henon maps ($T_{x_t\to y_t}$: dotted, $T_{y_t\to x_t}$: solid).

$s>8.0$ ($f<0.058$ Hz), however, the direction of the causal interaction is inverted, i.e., $T_{c,H\to B} < T_{c,B\to H}$. Such inversion of information flow is a documented phenomenon [21], but to our knowledge no conclusive evidence exists as to why it occurs. Our result may point to a hidden low frequency mechanism ($f<0.06$ Hz) underlying the causal influence that may be worthwhile investigating further.

Another interesting result is that $T_{c,B\to H}\approx 0$ in the scale interval [1.21, 3.15] (0.15–0.39 Hz); whereas $T_{c,H\to B}=0$ in the interval [1.46, 2.4] (0.19–0.32 Hz) and is <0.01 for 1.2

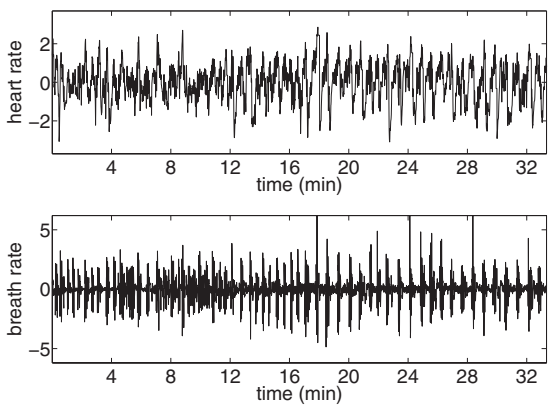

FIG. 5. Simultaneously recorded time series of heart rate (top) and lung pressure (breath rate) (bottom). The data corresponds to samples no. 305–4305 of the Santa Fe data set B archived in PhysioNet databank [19]. The sampling rate is $f_s=2$ Hz. Both traces are normalized to zero mean and unit variance.

$<s<4.0$ (which hints at a weak coupling). This finding is consistent with the known fact that the high frequency component of HRV (HF, 0.15–0.45 Hz) is synchronous with the respiratory cycle and is identical to the RSA (which is due to the parasympathetic activity). Notable are also the high values of $T_{c,B\to H}$ and $T_{c,H\to B}$ in the scale interval [5.5, 11.5] (0.0418–0.0625 Hz) which corresponds to the lower part of the low frequency component of the HRV spectrum (LF, 0.04–0.15 Hz). Note that all our findings are plausible under the assumption that the processes underlying heart rate and breath rate dynamics mutually causally influence each other (that is, both can either act as driver or as response). To conclude this forth application scenario, we point out that the proposed multiscale approach is rather specific and reasonably sensitive even when applied to a realistic data set, without requiring too much parameter tuning.

C. Robot data

The goal of our last application is to identify and map information flow in a simple robot (composed of one sensor and one actuator; henceforth the "controlled system") coupled to a chaotic map (henceforth the "controller"). We

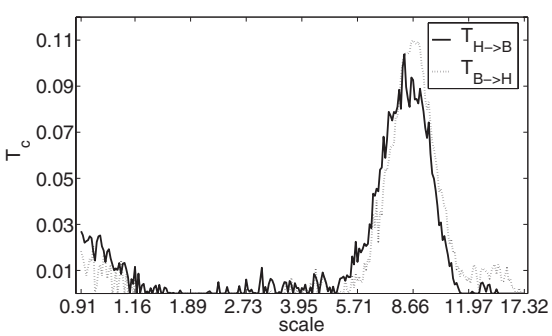

FIG. 6. Wavelet-based transfer entropy $T_{c,B\to H}$ (solid) and $T_{c,H\to B}$ (dashed). B=breath, H=heart. The parameters used are: $f_s=2$ Hz, $s_0=0.9$, $V=60$, $n=256$, and $\omega_0=6$.

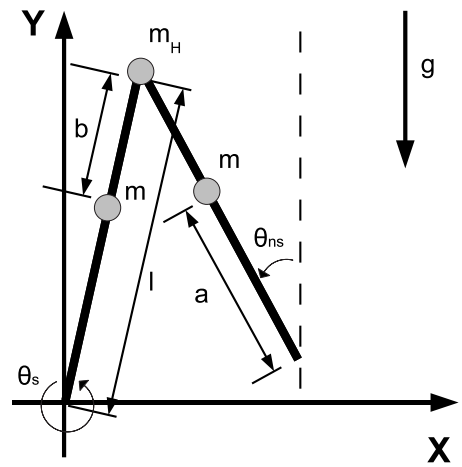

FIG. 7. Lumped model of compass-gait biped robot. The mass of each leg is m; the mass of the hip is m_H. The length of each leg is l (divided into a segment a and a segment b). θ_s and θ_{ns} are the angles formed by the supporting (stance) and the nonsupporting (swing) leg with the vertical (counterclockwise positive). The interleg angle is $\theta_{ns}-\theta_s$ and g is the gravity vector.

opted for a chaotic controller not only because it represents one of the most exciting theoretical outcomes of contemporary dynamical systems theory [22,23], but also because such a controller might provide an interesting alternative to conventional implementations. We emphasize that the goal of the present paper is not to discuss the chaos-based control framework which will be explained in depth in another paper.

Specifically, we collected field data using a two-dimensional two-link model kinematically equivalent to a double pendulum and vaguely resembling a human walking on level ground (Fig. 7). To simplify our analysis we restricted the motion of the biped to the sagittal plane, forcing the robot to behave like a compass (hence the name compass-gait biped robot). The biped has two rigid (kneeless) legs of point mass m connected by a frictionless hinge joint at the hip. A third point mass m_H representing the mass of the upper body is concentrated at the same joint. We assume that when the nonsupporting foot hits the ground its velocity jumps instantaneously to zero. That foot then becomes the supporting leg and stays on the ground acting like a hinge until the other foot touches the ground. The impact of the foot with the ground is modeled as an inelastic (no-slip, no-bounce) collision (see below)—energy is lost at each impact and in order for the system to keep walking (without falling), energy must be supplied externally.

The configuration of the compass-gait biped is determined by

$$\boldsymbol{\theta}=[\theta_{ns},\theta_s]^T, \tag{19}$$

where θ_{ns} and θ_s are the angles spanned by the nonsupport (swing) and the support (stance) leg with the vertical, respectively, and $\dot{\theta}_{ns}$ and $\dot{\theta}_s$ are their temporal derivatives (see Fig. 7). The equation of motion governing the dynamic behavior of the biped is

TABLE I. Parameters used for the simulated compass walker.

m_H (kg)	m (kg)	a (m)	b (m)	$l=a+b$ (m)	g (m/s^2)
10.0	5.0	0.5	0.5	1.0	9.81

$$M(\boldsymbol{\theta})\ddot{\boldsymbol{\theta}} + C(\boldsymbol{\theta},\dot{\boldsymbol{\theta}})\dot{\boldsymbol{\theta}} + g(\boldsymbol{\theta}) = \boldsymbol{S}\tau \tag{20}$$

with

$$M(\boldsymbol{\theta}) = \begin{bmatrix} mb^2 & -mbl\cos(\theta_s-\theta_{ns}) \\ -mbl\cos(\theta_s-\theta_{ns}) & (m+m_H)l^2+ma^2 \end{bmatrix}, \tag{21}$$

$$C(\boldsymbol{\theta},\dot{\boldsymbol{\theta}}) = \begin{bmatrix} 0 & -mbl\sin(\theta_{ns}-\theta_s)\dot{\theta}_s \\ -mbl\sin(\theta_{ns}-\theta_s)\dot{\theta}_{ns} & 0 \end{bmatrix}, \tag{22}$$

$$g(\boldsymbol{\theta}) = g\begin{bmatrix} mb\sin\theta_{ns} \\ -(m_H l+ma+ml)\sin\theta_s \end{bmatrix}, \tag{23}$$

where $M(\boldsymbol{\theta})$ is the 2×2 inertia matrix of the robot, $C(\boldsymbol{\theta},\dot{\boldsymbol{\theta}})$ is a 2×2 matrix containing centrifugal and Coriolis terms, $g(\boldsymbol{\theta})$ is a 2×1 vector of gravitational torques, $\boldsymbol{S}=[-1,1]^T$ is a conversion vector, and τ is the torque applied to the hip joint. The values of the parameters used in the simulation can be found in Table I. The equations of motion were numerically integrated using a fourth order Runge-Kutta with integration step $\Delta t=5\times 10^{-3}$ s.

Equation (20) adequately expresses the dynamics of the biped when one foot is on the ground while the other leg is swinging. The impact of the swinging feet with the ground, however, requires some additional considerations. Here, to model such impact we make the following simplifying assumptions: (a) the impact is perfectly plastic (no-bounce condition); (b) the legs do not slip (no-slip condition); (c) support is instantaneously transferred from the supporting leg to the swing (nonsupporting) leg; and (d) the angular momentum is conserved. For further elaborations on this issue, refer to [24].

The controller applies torque to the hip at continuous time $t_k \in R_0^+$, with $t_{k+1} > t_k$, $t_0=0$; in all our simulations, $t_k = k\Delta t$. The control events can be described by the following equations:

$$\tau_n = \gamma_2 c_n,$$

$$c_n = (1-\gamma_1)(1-\alpha c_{n-1}^2) + \gamma_1 f(\boldsymbol{\theta}),$$

$$f(\boldsymbol{\theta}) = \dot{\theta}_s - \dot{\theta}_{ns}, \tag{24}$$

where the control constant γ_2 determines the coupling strength between the dynamics of the controller and the controlled system (via torque), the parameter γ_1 sets the influence of the feedback term onto the evolution equation of the logistic map $[x_n=(1-\alpha x_{n-1}^2)]$, and α is the bifurcation parameter. For all our simulations, $\alpha=1.95$, for which the logistic map exhibits chaotic dynamics.

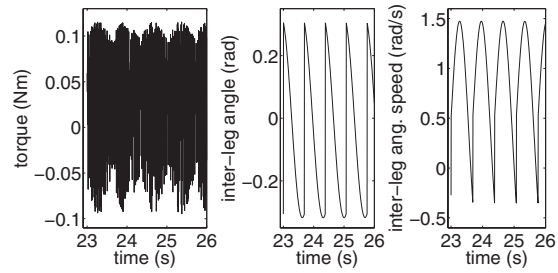

FIG. 8. Time series excerpt for closed-loop control scenario ($\gamma=0.11$). Torque applied to hip joint in Nm; interleg angle in rad; interleg angular speed in rad/s. The sampling frequency is 200 Hz.

We studied two control strategies: (1) *open-loop* control [$\gamma_1=0$ and $\gamma=\gamma_2>0$; the actuation dynamics and thus the torque τ_n applied to the hip joint are independent of the sensory feedback $f(\boldsymbol{\theta})$]; and (2) *closed-loop* control [$\gamma=\gamma_1=\gamma_2>0$; the state c_n of the chaotic map is affected by the controlled system via feedback, and conversely the controller affects the controlled system]. In both control scenarios, we varied the coupling parameter γ in the interval [0.03, 0.33] with increments of 0.003 and measured the wavelet-based transfer entropy between the torque τ (controlled variable) and the interleg angular velocity $\dot{\theta}_s - \dot{\theta}_{ns}$ (the measured variable). All runs had a maximum length of 15 000 time steps (equivalent to 75 s) and a minimum length of 1200 time steps (equivalent to 6 s). If the biped fell before the 6 s limit, the trial was labeled as unsuccessful, and the transfer entropy was set to zero. For every value of γ, we performed 30 runs with $\boldsymbol{\theta}(t=0)$ and $\dot{\boldsymbol{\theta}}(t=0)$ uniformly sampled from $[-0.4,0.4]\times[-0.4,0.4]$. An excerpt from a typical robot data set for $\gamma=\gamma_1=\gamma_2=0.11$ is reproduced in Fig. 8. We note that other control scenarios would have been possible, e.g., with asymmetric coupling between controller and controlled systems, that is, $\gamma_1<\gamma_2$ or $\gamma_1>\gamma_2$. For the sake of simplicity, however, we studied only open-loop control and "symmetric" closed-loop control.

Figures 9–12 show the results of our analysis ($\omega_0=5$, $s_0=1$, $V=10$, and $n=256$ scales). We can make a few observations. First off, in the closed-loop case the information transfer from the controlled system (body) to the controller (chaotic "neural" system) is clearly localized in frequency (Fig. 12). In other words, feedback of "meaningful" sensory information occurs at specific time scales. In the figure two peaks are discernible, one located in the scale interval [45, 60], the other in the interval [100, 140] (roughly equivalent to the frequency intervals 2.7–3.6 Hz and 1.1–1.8 Hz, respectively). Such localization is clearly a result of the feedback loop and is absent in the open-loop case. The time scale of the natural dynamics of the biped lies in the same range. For example, in Fig. 8 the stride frequency is 1.47 Hz. A second observation is that in the closed-loop case there is an evident asymmetry between information flowing into the controlled system $T_{c,b\to n}$ (peak: 0.445 bit) and information flow out of it $T_{c,n\to b}$ (peak: 0.021 bit)—despite symmetry in the coupling

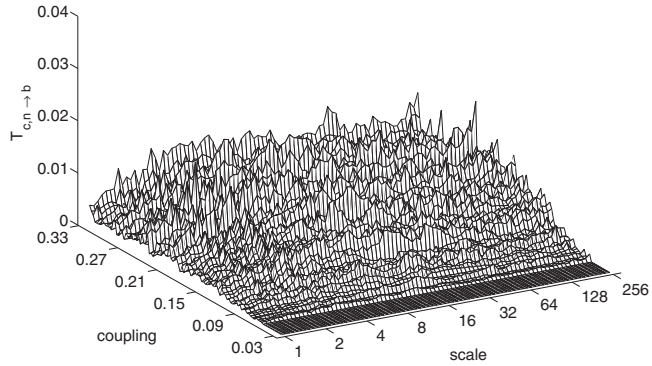

FIG. 9. Wavelet-based transfer entropy for open-loop system. Displayed is $T_{c,n\to b}$, where "n" indicates the controller (neural system) and "b" indicates the controlled system (body).

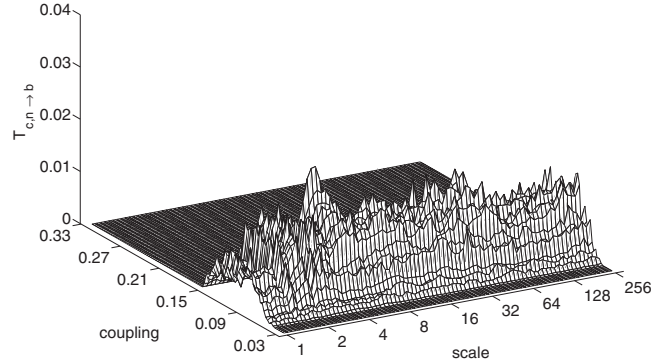

FIG. 11. Wavelet-based transfer entropy for closed-loop system. Displayed is $T_{c,n\to b}$, where "n" indicates the controller (neural system) and "b" indicates the controlled system (body).

($\gamma_1=\gamma_2$). One possible explanation is that for "good" values of γ, the self-stabilizing properties of the biped robot allow it to operate at a stable pseudoperiodic limit cycle. In a sense, the natural dynamics of the coupled system and the physical interaction induce essential information structure in the sensory feedback. This hypothesis is validated by Figs. 10 and 12 which show that the information flow is concentrated at lower frequencies. In both control scenarios, the controller affects the controlled system over a larger range of frequencies than vice versa (see Figs. 9 and 10)—as it ought to be, given that the controller is a chaotic system.

We further observe that a small amount of control information can have a large effect. In the closed-loop case, for instance, the flow of information from controller to controlled system $T_{c,n\to b}$ is one order of magnitude smaller than the flow from controlled system to controller $T_{c,b\to n}$. This result might be relevant from a design point of view. By focusing more on the natural dynamics of the controlled system (i.e., its body dynamics) and less on control (that is, strong coupling between controlled system and controlled), it might be possible to design more energy-efficient robotic systems (e.g., [25,26]). For instance, the positive mechanical work per cycle in the case of the symmetric closed-loop controller for $\gamma=0.11$ is approximately $E^{+}=29$ mJ/cycle (due to the instantaneous inelastic collisions of the feet with the ground energy is lost); whereas for the open-loop controller $E^{+}=47$ mJ/cycle. These are average values obtained over ten successful runs. A last observation concerns the open-loop case. For $\gamma<0.4$, the causal influence of the controller on the controlled system is small, and the self-stabilizing properties of the latter are sufficient for walking to be rather stable (the robot walks stably for more than 10 s) (see Fig. 13). For $\gamma>0.4$, however, such influence is too strong and walking does not occur. The robot falls often and on the average walks stably only for less than 7.5 s (data not shown).

IV. SUMMARY AND DISCUSSION

Systems in the real world are typically regulated by a complex web of dependencies and couplings exchanging information at multiple spatial and temporal scales. In the field of experimental signal analysis, a fundamental problem is to tell if a given set of scalar observations originates from interacting or noninteracting systems. In this paper we have proposed a wavelet-based extension of transfer entropy to analyze the causal dependencies between bivariate time series across multiple time scales. Using five different data sets collected from different dynamical systems, we have shown

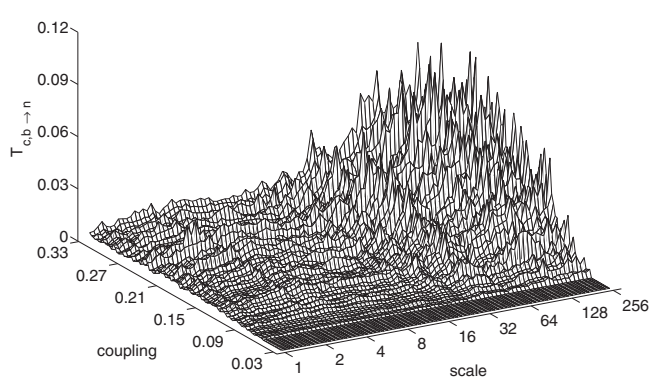

FIG. 10. Wavelet-based transfer entropy for open-loop system. Displayed is $T_{c,b\to n}$, where "b" indicates the controlled system (body) and "n" indicates the controller (neural system).

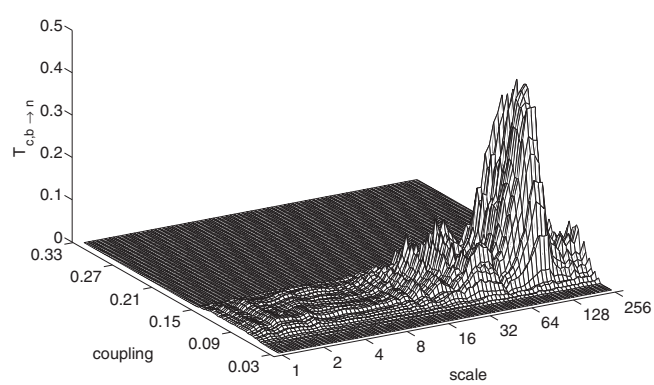

FIG. 12. Wavelet-based transfer entropy for closed-loop system. Displayed is $T_{c,b\to n}$, where "b" indicates the controlled system (body) and "n" indicates the controller (neural system).

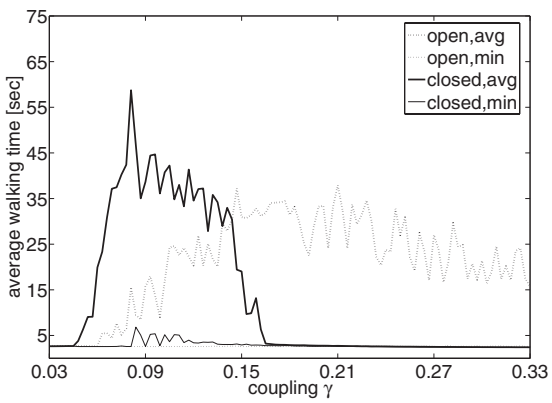

FIG. 13. Average walking time (before falling) of the compass-gait biped. On the average open-loop control leads to shorter walking times (due to its inherent instability).

that our approach yields consistent findings when applied to assess multiscale information flow between artificial, physiological, and robotic time series. Our experimental results are consistent with previous work confirming the effectiveness and validity of our approach. Our method is furthermore capable of detecting scale-dependent causal dependencies in a physiological data set undetectable with previous methods and provides an innovative view on information flow in a simulated chaos-based biped walker. Although at small scales (high frequencies) our approach delivers qualitatively similar results to other techniques [2,3,20,27], it yields insights about potential interactions at higher scales (lower frequencies). Our wavelet-based method appears to yield a meaningful approach to capture the coarse-to-fine "causal" structure of the interaction between coupled dynamical systems.

One particular problem affecting the current implementation is the finite sample size effect which grows fast as the number of data points is reduced. The method proposed here requires an adequate amount of data in order to provide sufficient statistical accuracy to estimate the transfer entropy. Throughout the paper, it was made sure that $N > N_{\min} \approx 3/r^3$ [11] (e.g., for the physiological time series $r=0.16$ and $N=4000$ samples, whereas $N_{\min}=733$). The finite sample effect is a known problem, and could be addressed either by replacing transfer entropy with another measure of causality [21], or by choosing a better kernel density estimator [28]. Alternatively, one could replace the kernel density estimator by some adaptive binning strategy, such as the one proposed in [21]. A related issue, particularly relevant in the case of a low number of data, is how to assess the significance of the detected multiscale information transfer. Here we built surrogate time series by shuffling (randomly reordering) the sequence of data points. In this way, we destroyed the correlations among the data points while preserving the statistical properties of the distribution, particularly, the first and second order moments. A similar technique was used in [18].

Another limitation of the proposed method is its bivariate nature. Some recent studies have emphasized the importance of a multivariate approach [21,29]. Moreover, many multi scale approaches to assess the complexity or the information flow between interconnected parts of a complex system have been suggested (e.g., [30–32]). Typically, multivariate nonlinear interdependencies between subsystems at multiple scales are analyzed by resampling the original time series at various scales yielding a collection of data possessing different coarse-graining from which various measures can be calculated (e.g., entropy, mutual information). For instance, in [33] it was shown that the heart-beat time series of healthy people asymptotically approaches a constant value of entropy as the measurement scale is increased. On the other hand, the analysis demonstrates that in patients suffering from "atrial fibrillation" or "congestive heart failure" the entropy as a function of time scale shows a consistently different pattern. The usage of such tools for diagnostic purposes is invaluable. Especially because the relevant time scales of such measured time series, such as the cardiac cycle, are not known *a priori*. Furthermore, they can be modified by autonomic blockades and under pathophysiological conditions. The approach introduced in this paper can be readily extended to incorporate multivariate techniques. One could, for instance, replace the original time series of any multivariate technique by the wavelet coefficients and then perform the analysis in scale space. Or, one could achieve the same result by appropriately conditioning the transition probabilities in Eq. (4) [7,21].

In this paper, we have considered the information transfer (causal structure) on the same scale. It might be worthwhile considering also information flow between different scales [34]. This could lead to meaningful insights into the dynamics of many biological systems in which processes working at dissimilar time scales coexist, e.g., the time scale of neural activity and the time scale of actions in the real world [30]. Such processes describe, in a sense, the informational embedding of organisms within their ecological niches at multiple time scales defining continuous and dynamic coupling between sensory, neural, and motor variables [35]. The comparison of the relative influence such variables exert on each other helps extract patterns of interaction between the network's elements that may support biological information processing. The measure introduced in this paper, which by using transfer entropy allows capturing directed exchanges of information (information flow) between sensory and motor variables in a physically embedded system, might prove useful for mapping the intrinsic dynamics of such sensorimotor networks.

To conclude, we suggest that the combination of wavelet transform and transfer entropy (or any other time series based measure to detect information transfer or couplings) can help shed light on the hidden causal dependencies in the coarse structure of the data from a wide range of complex systems, which a sample-based approach working at one single time scale is not able to disclose.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for their valuable comments on the draft of the manuscript.

[1] J. A. Vastano and H. L. Swinney, Phys. Rev. Lett. **60**, 1773 (1988).
[2] N. Ancona, D. Marinazzo, and S. Stramaglia, Phys. Rev. E **70**, 056221 (2004).
[3] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. E **73**, 066216 (2006).
[4] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, Phys. Lett. A **324**, 26 (2004).
[5] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, Physica D **134**, 419 (1999).
[6] U. Feldmann and J. Bhattacharya, Int. J. Bifurcation Chaos **14**, 505 (2004).
[7] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[8] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[9] M. Takens, in *Lecture Notes in Mathematics: Dynamical Systems and Turbulence*, edited by D. Rand and L. Young (Springer-Verlag, Berlin, 1981), Vol. 898, pp. 366–381.
[10] B. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986).
[11] M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu, Int. J. Bifurcation Chaos **17**, 903 (2007).
[12] R. Marschinski and H. Kantz, Eur. Phys. J. B **30**, 275 (2002).
[13] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, PA, 1992).
[14] S. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, San Diego, 1999).
[15] C. Torrence and G. P. Compo, Bull. Am. Meteorol. Soc. **79**, 61 (1998).
[16] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. (Cambridge University Press, Cambridge, UK, 2002).
[17] T. Schreiber and A. Schmitz, Physica D **142**, 346 (2000).
[18] R. G. Andrzejak, A. Ledberg, and G. Deco, New J. Phys. **8**, 6 (2006).
[19] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, Circulation **101**, e215 (2000).
[20] J. Bhattacharya, E. Pereda, and H. Petsche, IEEE Trans. Syst., Man, Cybern., Part B: Cybern. **33**, 85 (2003).
[21] P. F. Verdes, Phys. Rev. E **72**, 026222 (2005).
[22] E. Ott and M. Spano, Phys. Today **48**(5), 34 (1995).
[23] J. M. Gonzalez-Miranda, *Synchronization and Control of Chaos* (Imperial College Press, London, 2004).
[24] T. McGeer, Int. J. Robot. Res. **9**, 62 (1990).
[25] S. Collins, A. Ruina, R. Tedrake, and M. Wisse, Science **307**, 1082 (2005).
[26] K. Matsushita, M. Lungarella, C. Paul, and H. Yokoi, in *Proceedings of the 20th International Conference on Robotics and Automation* (2005), pp. 2020–2025.
[27] L. Angelini, R. Maestri, D. Marinazzo, L. Nitti, M. Pellicoro, G. D. Pinna, S. Stramaglia, and S. A. Tupputi (unpublished).
[28] Y.-I. Moon, B. Rajagopalan, and U. Lall, Phys. Rev. E **52**, 2318 (1995).
[29] K. J. Blinowska, R. Kus, and M. Kaminski, Phys. Rev. E **70**, 050902(R) (2004).
[30] M. Breakspear and C. Stam, Philos. Trans. R. Soc. London, Ser. B **360**, 865 (2005).
[31] M. Costa, A. L. Goldberger, and C. K. Peng, Phys. Rev. E **71**, 021906 (2005).
[32] S. Thurner, M. C. Feurstein, and M. C. Teich, Phys. Rev. Lett. **80**, 1544 (1998).
[33] M. Costa, A. L. Goldberger, and C. K. Peng, Phys. Rev. Lett. **89**, 068102 (2002).
[34] M. Materassi, A. Wernik, and E. Yordanova, Nonlinear Processes Geophys. **14**, 153 (2007).
[35] M. Lungarella and O. Sporns, PLOS Comput. Biol. **2**, 1301 (2006).
[36] The general nature of our results does not critically depend on the exact choice of the mother wavelet.