34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Compensating for Instantaneous Signal Mixing in Transfer Entropy Analysis of Neurobiological Time Series

Luca Faes, *Member, IEEE*, Silvia Erla, and Giandomenico Nollo, *Member, IEEE*

Abstract— **The transfer entropy (TE) has recently emerged as a nonlinear model-free tool, framed in information theory, to detect directed interactions in coupled processes. Unfortunately, when applied to neurobiological time series TE is biased by signal cross-talk due to volume conduction. To compensate for this bias, in this study we introduce a modified TE measure which accounts for possible instantaneous effects between the analyzed time series. The new measure, denoted as compensated TE (cTE), is tested on simulated time series reproducing conditions typical of neuroscience applications, and on real magnetoencephalographic (MEG) multi-trial data measured during a visuo-tactile cognitive experiment. Simulations show that cTE performs similarly to TE in the absence of signal cross-talk, and prevents false positive detection of information transfer in the case of instantaneous mixing of uncoupled signals. When applied to MEG data, cTE detects significant information flow from the visual cortex to the somatosensory area during task execution, suggesting the activation of mechanisms of multisensory integration.**

I. Introduction

The transfer entropy (TE) is an information theoretic measure of directed information transfer between interacting processes. Since its first introduction by Schreiber [1], TE has been recognized as a powerful tool for detecting causal interactions in time series data measured from coupled systems, as it offers an approach that is free of an explicit model of the studied dynamics and is sensitive to both linear and nonlinear interdependencies. The popularity of this tool has grown even more in recent years, thanks to the development of data-efficient estimation procedures which favored reliable TE computation from neurobiological data such as electroencephalographic (EEG) or magnetoencephalographic (MEG) time series [2-6].

One major problem in studying interactions in non-invasive multichannel neurobiological recordings such as EEG or MEG are the artifacts of volume conduction. Volume conduction is a consequence of the simultaneous mapping of single sources of brain activity, which are located inside the brain, onto several recording sensors, which are located on the scalp. The instantaneous mixing of unmeasured sources results in a nontrivial interference between the measured sensor data which unavoidably affects connectivity analyses performed at the sensor level [7,8]. In TE analysis, volume conduction artifacts can result in false positive detection of information transfer between pairs of channels both affected by the same cortical source having an internal memory structure [6].

L. Faes, G. Nollo and S. Erla are with the Dept. of Physics and BIOtech, University of Trento, Mattarello (TN), Italy (corresponding author e-mail: luca.faes@unitn.it).

The aim of this study is to introduce a modified TE measure which is able to compensate for the spurious detection of information transfer between signals due to instantaneously shared background activity. The novel TE measure is proposed in combination with an approach for the efficient estimation of information transfer from short time series data [4], validated in comparison with the traditional TE on simulated systems reproducing conditions typical of neuroscience applications, and finally applied on MEG data recorded during a visuomotor integration task.

II. Methods

A. Transfer Entropy

Consider two stochastic processes x and y representing the evolution over time of the physical systems X and Y. Let x_t and y_t be the stochastic variables which describe the state visited by X and Y at the time t, and $x_{n:t}$ be the vector composed of all samples of x from time n up to time t. Then, the transfer entropy (TE) from X and Y is defined as [1]

$$TE_{X\rightarrow Y} = H(y_t|y_{1:t-1}) - H(y_t|y_{1:t-1}, x_{1:t-1}), \qquad (1)$$

where $H(\cdot)$ denotes Shannon entropy, which represents the uncertainty associated with any measurement a of a vector random variable $\boldsymbol{a}$, $H(\boldsymbol{a})= -\sum_a p(a)\log p(a)$, and $H(\cdot|\cdot)$ denotes conditional entropy (CE), which represents the uncertainty that remains about $\boldsymbol{a}$ when $\boldsymbol{b}$ is known, $H(\boldsymbol{a}|\boldsymbol{b})=H(\boldsymbol{a},\boldsymbol{b})-H(\boldsymbol{b})$. The TE defined in (1) quantifies the information transfer from X and Y as the amount of information carried by the most recent sample of the destination process y (i.e., y_t) which is not contained in the past of the source process x (i.e., $x_{1:t-1}$; here the origin of time is conventionally set at t=1).

B. Compensated Transfer Entropy

In the case in which the process x has an internal memory structure (i.e., x_t is partly explained by $x_{1:t-1}$) and is instantaneously correlated with the process y (i.e., y_t is partly explained by x_t), the TE defined in (1) can take significant positive values even if $x_{1:t-1}$ is not useful to explain y_t. In this situation, which may reflect e.g. the instantaneous mixing of a common signal into the two observed processes, information transfer from X to Y may be detected using the TE even when the two systems are uncoupled. To counteract this problem, we propose to incorporate instantaneous effects in both the CE terms into which TE computation is factored. Accordingly, we define the compensated TE (cTE) as

$$TE^c_{X\rightarrow Y} = H(y_t|y_{1:t-1}, x_t) - H(y_t|y_{1:t-1}, x_{1:t}). \qquad (2)$$

With this definition, theoretical values of cTE are nonzero only when the knowledge of $x_{1:t-1}$ is useful to explain y_t above

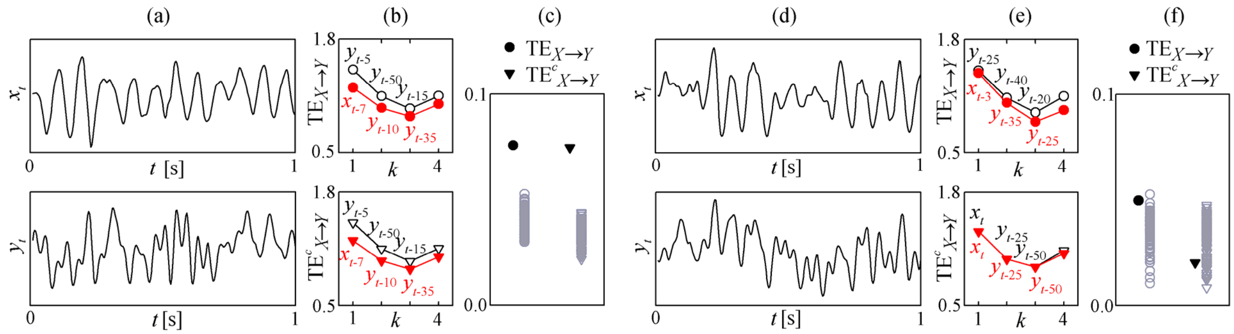

Figure 1. Example of computation of TE and cTE over realizations of the simulation in (4) with parameters C=0.4, ε=0 (a,b,c) and C=0, ε=0.4 (d,e,f). (a,d) single realizations of the simulated processes; (b,e) estimation of $H(y_t|y_{1:t-1})$ (black circles), $H(y_t|y_{1:t-1},x_{1:t-1})$ (red circles), $H(y_t|y_{1:t-1},x_t)$ (black triangles) and $H(y_t|y_{1:t-1},x_{1:t})$ (red triangles) using nonuniform embedding, with indication of the candidate terms selected at each step k of the sequential procedure; (c,f) median values of TE (circles) and cTE (triangles) estimated over 50 simulation realizations (black solid) and over 100 permutations of the same realizations in which the signals for X and Y were randomly shuffled (gray open).

and beyond the knowledge of $y_{1:t-1}$ and x_t alone. Note that in the absence of instantaneous correlation between x_t and y_t the cTE reduces to the traditional TE defined in (1).

C. Transfer Entropy Estimation

The practical computation of TE and cTE is based on reconstructing the state space of the systems X and Y from the available time series data. The typical approach to state space reconstruction is based on uniform time delay embedding, so that, e.g., $x_{1:t-1}$ is approximated with the delay vector $(x_{t-u}, x_{t-u-\tau}, ..., x_{t-u-(d-1)\tau})$. This procedure requires setting the prediction time u, as well as the embedding delay τ and dimension d, which is a crucial but not easy task to perform [6]. To overcome the issues of arbitrariness and redundancy associated with uniform embedding, in this study we used the approach proposed in [4].

The approach follows a procedure for nonuniform embedding whereby delay vectors are formed in a sequential way selecting progressively the samples that contribute most to the description of the observed dynamics. These samples are taken from a set of initial candidates which includes the past (and, when appropriate, present) states of X and Y, combined as indicated in (1) and (2) for the estimation of the various CE terms; for instance, the set of candidates for estimating $H(y_t|y_{1:t-1}, x_{1:t})$ in (2) will be the vector of $2L$+1 terms given by $(y_{t-\tau}, ... , y_{t-L\tau}, x_t, x_{t-u}, x_{t-u-\tau}, ... , x_{t-u-(L-1)\tau})$. Starting with an empty embedding vector, at each step the sequential procedure tests all candidates and then includes into the vector the candidate which minimizes the CE estimate. The procedure stops when a minimum of the CE is reached, and this minimum is used as in (1) and (2) for TE and cTE computation. As to CE estimation, the procedure exploits the corrected estimator proposed in [9], which is based on uniform quantization of the time series and compensation of the CE bias through introduction of a corrective term which penalizes isolated points within state spaces of increasing dimensions. While illustrative examples are shown in Fig. 1, we refer to [4] for a detailed description of the overall procedure.

III. Simulated Data

To test the ability of TE and cTE to measure information transfer under situations relevant to neuroscience applications, we considered the simulated systems X and Y described by the stochastic processes x' and y' as

$$\begin{aligned} x'_n &= a_1 x'_{n-1} + a_2 x'_{n-2} + u_n \\ y'_n &= \sum_{k=1}^{4} b_k y'_{n-k} + C x'_{n-1} + w_n \end{aligned} , \quad (3)$$

where u and w are uncorrelated white noise processes with zero mean and unit variance (n is the discrete time index). The parameters in (3) were set at a_1=1.385, a_2=-0.9025, b_1=1.4266, b_2=-1.465, b_3=1.2875, b_4=-0.5077, while the parameter C was let free to vary from 0 to 1 to mimic different degrees of unidirectional coupling from x' to y'. With this choice and assuming an initial sampling rate of 100 Hz, (3) generates oscillations at ~12 Hz for x', and at ~5 Hz and ~25 Hz for y', simulating respectively the presence of alpha, delta/theta and beta brain waves.

After generating realizations of (3) of 1 s duration (i.e. 100 samples length), we upsampled the series to get the series x'_t and y'_t measured at a simulated sampling frequency of 300 Hz. Then, instantaneous mixing of x' into the two measured processes x and y was obtained as

$$\begin{aligned} x_t &= x'_t \\ y_t &= \varepsilon x'_t + (1-\varepsilon) y'_t \end{aligned} , \quad (4)$$

where the parameter ε sets the amount of signal cross-talk. From these simulated signals, estimation of TE and cTE was performed setting the propagation time at u=4 ms and the delay embedding τ at the autocorrelation decay time (*act*) of each time series x_t and y_t, including L=10 terms in each set of initial candidates for the two systems, and using six levels for uniform quantization of the series in CE estimation. The analysis was repeated for 50 realizations (trials) of (4) for different selected combinations of C and ε. The statistical significance of the information transfer measured over the 50 trials was assessed using a permutation test in which the surrogate distribution of TE or cTE in the absence of

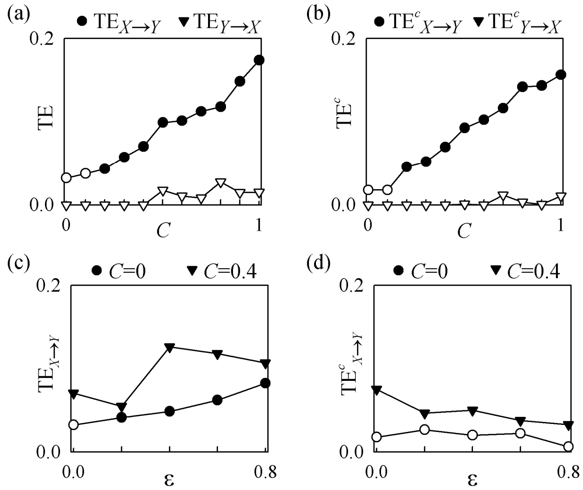

Figure 2. Median of TE (a) and of cTE (b) over 50 realizations of (4) computed from X to Y (circles) and from Y to X (triangles) as a function of the coupling strength C, with absence of signal cross-talk (ε=0); median of TE (c) and of cTE (d) computed from X to Y as a function of the amount of signal cross-talk ε, with absence (C=0, circles) and presence (C=0.4, triangles) of unidirectional coupling. In all plots, filled circles represent significant TE or cTE values assessed by means of the permutation test.

coupling was reproduced through repeated random pairing of the signals from X and Y across trials.

An example of a single realization of (4) generated with significant coupling but absence of instantaneous mixing (C=0.4, ε=0) is reported in Fig. 1a. The corresponding analysis of information transfer from X to Y, depicted in Fig. 1b, shows that TE and cTE take the same value in this situation. Indeed, in both cases the nonuniform embedding procedure selects progressively the terms y_{t-5}, y_{t-50} and y_{t-15} when estimating the CE of Y conditioned to its past (black), and the terms x_{t-7}, y_{t-10} and y_{t-35} when estimating the CE of Y conditioned to the past of X and Y (red). In all cases, the minimum CE is found at the step k=3 of the sequential procedure, yielding $TE_{X\to Y}=TE^c_{X\to Y}$=0.076 for this example. The same analysis extended to all 50 simulation trials confirms that TE and cTE detect the same information flow from X to Y, which is statistically significant as both TE and cTE median values lie clearly outside their corresponding distribution estimated by the permutation test (Fig. 1c). On the contrary, when considering realizations of (4) generated with absent coupling but significant mixing (C=0, ε=0.4) we see that only cTE correctly detects the absence of coupling, while CE provides a misleading indication of information transferred from X to Y. The analysis performed on the realization in Fig. 1d shows indeed that the past of X enters the embedding vector with the term x_{t-3} at the second CE estimation (Fig. 1e, up), yielding $TE_{X\to Y}$=0.049, while the instantaneous term x_t enters both embedding vectors and this prevents the selection of any past term from X (Fig. 1e, down), yielding $TE^c_{X\to Y}$=0. The analysis performed over all trials shows that the median TE is significant as it is located at the upper bound of its surrogate distribution, while the median cTE is clearly non significant (Fig. 1f).

The overall analysis summarized in Fig. 2 shows that TE and cTE perform very similarly in the absence of instantaneous mixing (ε=0, Fig. 2a,b), as both measures increase with the coupling strength C and are statistically significant for each C>0.1 when computed from X to Y, and are very low and never significant when computed over the uncoupled direction from Y to X. The difference between the two measures becomes apparent in the presence of significant signal cross-talk: while the CE gives a misleading indication of coupling from X to Y for uncoupled systems with instantaneous mixing (C=0, ε>0, Fig. 2c), the cTE correctly detects coupling only when C=0.4 and is never significant when C=0 regardless of the amount ε of signal cross-talk (Fig. 2d).

IV. MEG Experiment

The analyzed MEG signals were taken from a database of neurobiological recordings acquired during a visuo-tactile cognitive experiment [10]. Briefly, a healthy volunteer underwent a recording session in which simultaneous visual and tactile stimuli were repeatedly presented (60 trials). At each trial, geometric patterns resembling letters of the Braille code were both shown on a monitor and embossed on a tablet, and the subject had to perceive whether the pattern seen on the screen was the same of that touched on the tablet. The MEG signals (VSM whole head system) recorded during two consecutive time frames of 1 s, just before (rest window) and just after (task window) the presentation of the combined stimuli, were made available at a sampling frequency of 293 Hz.

To focus on the relevant information, two representative signals located in the somatosensory cortex (system X) and in the visual cortex (system Y) were considered for the analysis. At each trial, sensor selection was performed through a suitable event-related field analysis looking for the scalp location at which the signal magnitude was maximized in response to pure-visual or pure-tactile stimulation [10].The preprocessing consisted in FFT band-pass filtering (2-45 Hz) and removal of the event-related field from each task window by subtraction of the average response (over the 60 trials). An example of the analyzed signals is shown in Fig. 3a,b. The cTE analysis, performed with the same parameter setting of simulations (u=4 ms, τ=act, L=10), shows absent or very low amount of information transfer at rest (Fig. 3c), but reveals a non-negligible transfer from the visual to the somatosensory areas (measured by $TE^c_{Y\to X}$) during task (Fig. 3d). The overall analysis extended to the 60 trials shows that the information transfer is balanced but not statistically significant, according to the permutation test, in the rest condition (Fig. 3e), while it is markedly unbalanced with a prevalent and statistically significant transfer from visual to somatosensory areas during task execution (Fig. 3f).

Statistical analysis of the cTE distributions over trials, performed according to a Student t-test for paired data, indicated that $TE^c_{X\to Y}$ and $TE^c_{Y\to X}$ were balanced at rest (p=NS), while moving to the task window $TE^c_{Y\to X}$ increased significantly ($p<0.01$), becoming significantly higher than $TE^c_{X\to Y}$ ($p<0.001$). Thus, statistical analysis confirmed the prevalence of the information transfer from visual to somatosensory areas during execution of the combined visuo-motor task.

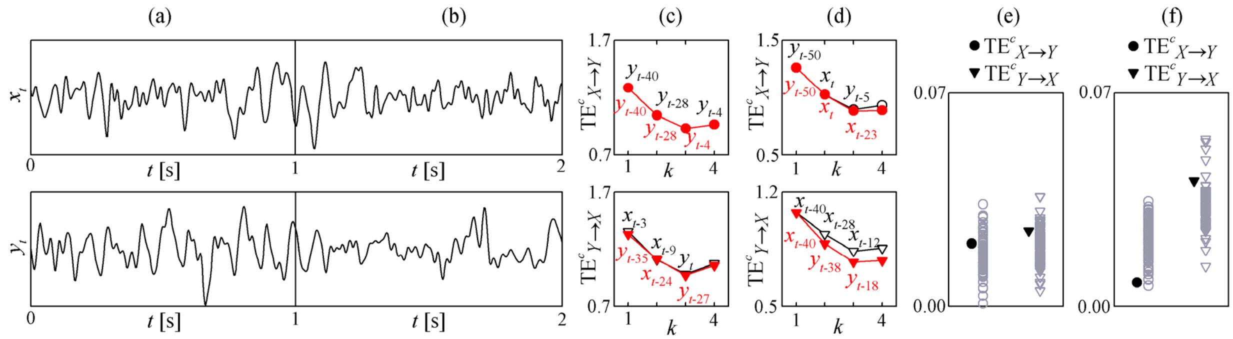

Figure 3. MEG signals recorded from the somatosensory cortex (x_t) and the visual cortex (y_t) for a representative trial before (a, rest window) and after (b, task window) stimulus presentation, and corresponding estimation of cTE over the two directions of interaction for the rest window (c) and the task window (d). The overall analysis performed for all 60 trials is reported in (e) for the rest window and in (f) for the task window. Plots and symbols are as in Fig. 1.

V. Discussion

Being model-free and able to capture both linear and nonlinear interactions, TE is a very flexible tool for the assessment of information transfer in coupled systems. Two main limitations regarding its application to neuroscience data are the difficulty of estimating entropies from short length datasets which hampers practical TE computation, and the confounding effects arising from the instantaneous mixing of unmeasured cortical sources which exposes TE to false positive detection of information transfer. The present study deals with these two problems, showing how they can be circumvented providing a tool that can reliably assess information transfer from pairs of neurobiological recordings. First, we combined TE with a data-efficient estimation procedure using nonuniform embedding and corrected CE [4]. We showed using simulations that, when the available data set has a trial structure, the proposed approach combined with proper statistical testing allows to detect information transfer for signals as short as 1 sec. Second, we faced the issue of instantaneous mixing allowing for the possibility of zero-lag effects in the computation of the two CE terms that enter the TE measure. Our simulations showed that the erroneous detection of information transfer for instantaneously mixed uncoupled signals may be prevented using this compensation. The proposed compensation is alternative to, and should be computationally more efficient than, the test of time-shifted data recently proposed to detect instantaneous mixing [3,6]. The application of the proposed approach to MEG sensor-level signals, where the problem of instantaneous mixing is inherent in the measurement method, suggests the feasibility of cTE for measuring information transfer across different brain regions during cognitive experiments. Our results, though certainly preliminary as obtained on a single subject, show that the considered matching paradigm may be able to evoke a response to combined visual and somatosensory stimulation, and that such a response may be quantified using the proposed cTE measure. Specifically, we hypothesize a specific role of the visual cortex in driving coherent activation of the somatosensory area during the execution of the combined visuo-tactile task, according to the involvement of multisensory integration mechanisms [11].

Acknowledgment

The authors would like to thank Christoph Braun and Carola Arfeller for providing magnetoencephalographic data from the MEG cente of the Eberhard-Karls-University of Tübingen, Germany.

References

[1] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 2000, no. 85, pp. 461-464, 2000.

[2] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Phys. Rep.*, vol. 441 pp. 1-46, 2007.

[3] M. Wibral, B. Rahm, M. Rieder, M. Lindner, R. Vicente, and J. Kaiser, "Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks," *Prog. Biophys. Mol. Bio.*, vol. 105, no. 1-2, pp. 80-97, Mar.2011.

[4] L. Faes, G. Nollo, and A. Porta, "Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique," *Phys. Rev. E*, vol. 83, no. 5-1, pp. 051112, May2011.

[5] V. A. Vakorin, O. A. Krakovska, and A. R. McIntosh, "Confounding effects of indirect connections on causality estimation," *J. Neurosci. Methods*, vol. 184, no. 1, pp. 152-160, Oct.2009.

[6] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy-a model-free measure of effective connectivity for the neurosciences," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 45-67, Feb.2011.

[7] L. Marzetti, C. Del Gratta, and G. Nolte, "Understanding brain connectivity from EEG data by identifying systems composed of interacting sources," *Neuroimage*, vol. 42, no. 1, pp. 87-98, Aug.2008.

[8] G. Nolte, A. Ziehe, V. V. Nikulin, A. Schlogl, N. Kramer, T. Brismar, and K. R. Muller, "Robustly estimating the flow direction of information in complex physical systems," *Phys. Rev. Lett.*, vol. 100, no. 23, pp. 234101, June2008.

[9] A. Porta, G. Baselli, F. Lombardi, N. Montano, A. Malliani, and S. Cerutti, "Conditional entropy approach for the evaluation of the coupling strength," *Biol. Cybern.*, vol. 81 pp. 119-129, 1999.

[10] S. Erla, C. Papadelis, L. Faes, C. Braun, and G. Nollo, "Studying brain visuo-tactile integration through cross-spectral analysis of human MEG recordings," in *MEDICON 2010, IFMBE Proceedings*, vol. 29, P. D. Bamidis and N. Pallikarakis, Eds., pp. 73-76, 2010.

[11] M. Bauer, "Multisensory integration: A functional role for inter-area synchronization?," *Curr. Biol.*, vol. 18, no. 16, pp. R709-R710, Aug.2008.