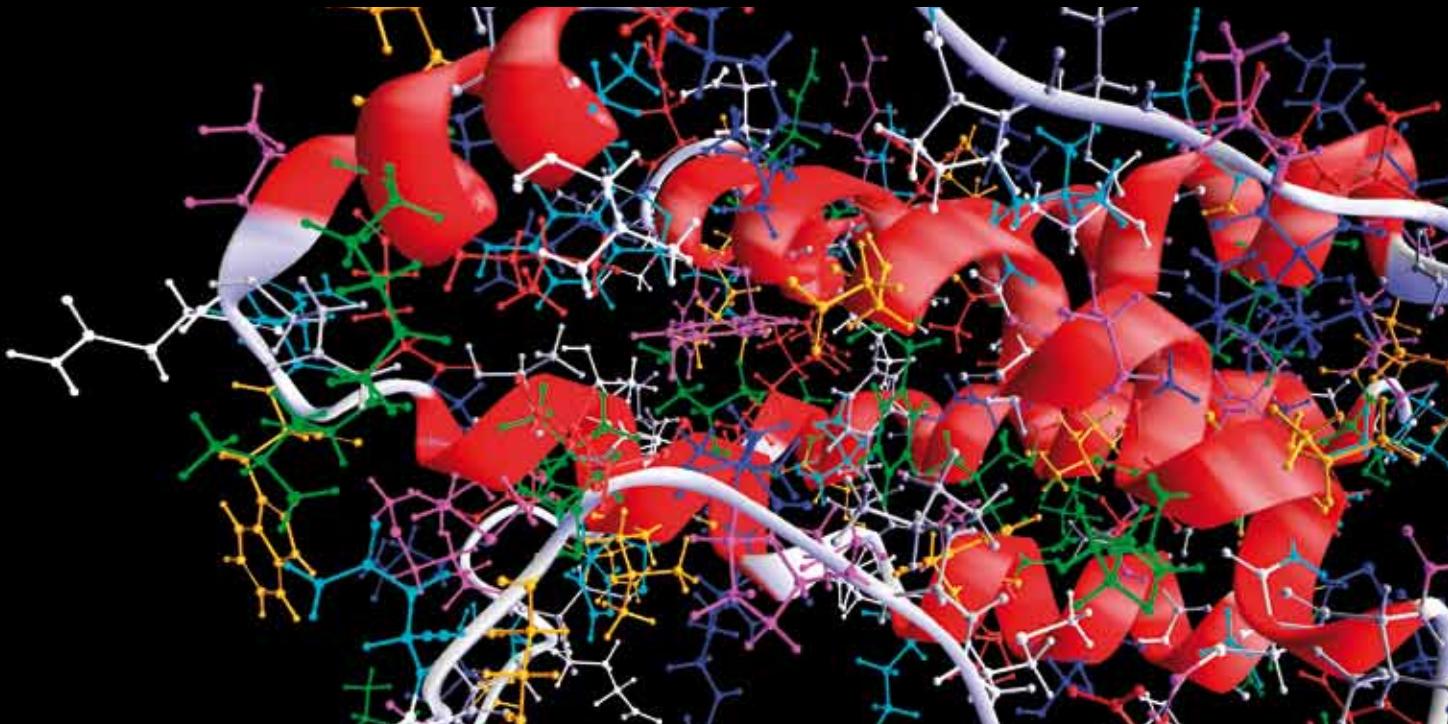


METHODOLOGICAL ADVANCES IN BRAIN CONNECTIVITY

GUEST EDITORS: LUCA FAES, RALPH G. ANDRZEJAK, MINGZHOU DING,
AND DIMITRIS KUGIUMTZIS





Methodological Advances in Brain Connectivity

Computational and Mathematical Methods in Medicine

Methodological Advances in Brain Connectivity

Guest Editors: Luca Faes, Ralph G. Andrzejak,
Mingzhou Ding, and Dimitris Kugiumtzis



Copyright © 2012 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "Computational and Mathematical Methods in Medicine." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Zvia Agur, Israel
Emil Alexov, USA
Gary C. An, USA
Georgios Archontis, Cyprus
Pascal Auffinger, France
Facundo Ballester, Spain
Dimos Baltas, Germany
Chris Bauch, Canada
Maxim Bazhenov, USA
Niko Beerenwinkel, Switzerland
Philip Biggin, UK
Michael Breakspear, Australia
Thierry Busso, France
Carlo Cattani, Italy
Bill Crum, UK
Timothy David, New Zealand
Gustavo Deco, Spain
Carmen Domene, UK
Wim Van Drongelen, USA
Frank Emmert-Streib, UK
Ricardo Femat, Mexico
Alfonso T. García-Sosa, Estonia
Kannan Gunasekaran, USA

Damien R. Hall, Japan
William F. Harris, South Africa
Vassily Hatzimanikatis, USA
Volkhard Helms, Germany
J.-H. S. Hofmeyr, South Africa
Seiya Imoto, Japan
Breddyn Jones, UK
Lawrence A. Kelley, UK
Lev Klebanov, Czech Republic
Ina Koch, Germany
David Liley, Australia
Quan Long, UK
Yoram Louzoun, Israel
Jianpeng Ma, USA
C.-M. C. Ma, USA
Reinoud Maex, France
Francois Major, Canada
Simeone Marino, USA
Ali Masoudi-Nejad, Iran
Ramit Mehr, Israel
Seth Michelson, USA
Michele Migliore, Italy
Karol Miller, Australia

Ernst Niebur, USA
Kazuhisa Nishizawa, Japan
Martin Nowak, USA
Markus Owen, UK
Hugo Palmans, UK
Lech S. Papiez, USA
Jean Pierre Rospars, France
David James Sherman, France
Sivalab Sivaloganathan, Canada
Stephen W. Smye, UK
Merryn H. Tawhai, New Zealand
Elisabeth Tillier, Canada
Nestor V. Torres, Spain
Anna Tramontano, Italy
Nelson J. Trujillo-Barreto, Cuba
Kutlu O. Ulgen, Turkey
Nagarajan Vaidehi, USA
Edelmira Valero, Spain
Jinliang Wang, UK
Jacek Waniewski, Poland
Guang Wu, China
Henggui Zhang, UK

Contents

Methodological Advances in Brain Connectivity, Luca Faes, Ralph G. Andrzejak, Mingzhou Ding, and Dimitris Kugiumtzis
Volume 2012, Article ID 492902, 2 pages

Fundamental Dynamical Modes Underlying Human Brain Synchronization, Catalina Alvarado-Rojas and Michel Le Van Quyen
Volume 2012, Article ID 912729, 8 pages

Localizing True Brain Interactions from EEG and MEG Data with Subspace Methods and Modified Beamformers, Forooz Shahbazi Avarvand, Arne Ewald, and Guido Nolte
Volume 2012, Article ID 402341, 11 pages

Causal Information Approach to Partial Conditioning in Multivariate Data Sets, D. Marinazzo, M. Pellicoro, and S. Stramaglia
Volume 2012, Article ID 303601, 8 pages

Quantification of Effective Connectivity in the Brain Using a Measure of Directed Information, Ying Liu and Selin Aviyente
Volume 2012, Article ID 635103, 16 pages

An Analytical Approach to Network Motif Detection in Samples of Networks with Pairwise Different Vertex Labels, Christoph Schmidt, Thomas Weiss, Christian Komusiewicz, Herbert Witte, and Lutz Leistritz
Volume 2012, Article ID 910380, 12 pages

Measuring Connectivity in Linear Multivariate Processes: Definitions, Interpretation, and Practical Analysis, Luca Faes, Silvia Erla, and Giandomenico Nollo
Volume 2012, Article ID 140513, 18 pages

Statistical Analysis of Single-Trial Granger Causality Spectra, Andrea Brovelli
Volume 2012, Article ID 697610, 10 pages

Enhancing the Signal of Corticomuscular Coherence, Cristiano Micheli and Christoph Braun
Volume 2012, Article ID 451938, 10 pages

Computational Models of Neuron-Astrocyte Interactions Lead to Improved Efficacy in the Performance of Neural Networks, Alberto Alvarellos-González, Alejandro Pazos, and Ana B. Porto-Pazos
Volume 2012, Article ID 476324, 10 pages

Source Activity Correlation Effects on LCMV Beamformers in a Realistic Measurement Environment, Paolo Belardinelli, Erick Ortiz, and Christoph Braun
Volume 2012, Article ID 190513, 8 pages

Inferring Functional Neural Connectivity with Phase Synchronization Analysis: A Review of Methodology, Junfeng Sun, Zhijun Li, and Shanbao Tong
Volume 2012, Article ID 239210, 13 pages

Editorial

Methodological Advances in Brain Connectivity

Luca Faes,¹ Ralph G. Andrzejak,² Mingzhou Ding,³ and Dimitris Kugiumtzis⁴

¹Department of Physics and BIOTech, University of Trento, 38123 Mattarello, Italy

²Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain

³J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611-6131, USA

⁴Department of Mathematical, Physical and Computational Sciences, Faculty of Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Correspondence should be addressed to Luca Faes, luca.faes@unitn.it

Received 5 August 2012; Accepted 5 August 2012

Copyright © 2012 Luca Faes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Determining how distinct neurons or brain regions are connected and communicate with each other is a crucial point in neuroscience, as it allows to investigate how the functional integration of specialized neural populations enables the emergence of coherent cognitive and behavioral states. The general concept of brain connectivity encompasses different aspects: structural connectivity is related to the description of anatomical pathways and synaptic connections; functional connectivity investigates statistical dependencies between spatially separated brain regions; effective connectivity refers to models aimed at elucidating driver-response relationships. The study of these different modes of brain connectivity is fundamental both for the investigation of the neurophysiological processes engaged in cognitive and perceptive processing and for the assessment of a variety of neurological disorders.

Over the last past few decades, the continuous advancement of multichannel data acquisition technologies has made it possible to collect neuroscience data at multiple levels of description, ranging from neural spikes to local field potentials and electroencephalography (EEG)-magnetoencephalography (MEG). This increasing data availability and the contemporaneous improvement of signal processing capabilities have contributed in placing increasing demands on methods for the quantitative assessment of neural interactions. Accordingly, a variety of methodological approaches have emerged for the estimation of brain connectivity from multivariate neurophysiological time series. These approaches build on different analysis frameworks which point out specific aspects of connectivity: linear parametric models are explicitly related to the frequency domain

representation of multivariate data, thus favoring the assessment of connectivity for specific brain rhythms; information theory provides tools that describe both linear and nonlinear interactions and are free from the shortcomings of model specification; phase synchronization analysis characterizes the relation between the phases of different brain units seen as coupled oscillators, allowing detection of synchronization regardless of the relation between signal amplitudes. All these frameworks provide measures able to reflect the various aspects of brain connectivity, such as undirected measures of association (e.g., coherence, mutual information, phase synchronization index) and directed measures of Granger causality (e.g., partial directed coherence, transfer entropy).

To get new and deeper insights into the complex dynamics of interacting brain regions, the existing brain connectivity measures need to be refined, extended, and complemented with other tools. Common issues to be addressed are estimation problems arising in the presence of noise contamination and nonstationarity, significance assessment, distinguishing direct from indirect causal effects, and aspects related to the difficulty of performing full multivariate analyses over short data sets. An important problem of EEG/MEG recordings is that the connectivity patterns estimated at the sensor level may strongly differ from those really existing between the underlying neural sources, as a consequence of a mixing effect known as volume conduction. While a common approach to deal with this issue is to estimate connectivity after application of an inverse method for source localization, methodological advances are needed to improve the localization accuracy

of ill-posed inverse problems. Moreover, once brain connectivity measures are computed, the formal representation of connectivity patterns in graph or matrix format leads to employ concepts of network analysis to interpret these patterns. Accordingly, the study of topological properties like clustering degree, modularity, and presence of network motifs is becoming increasingly popular for the investigation of the organizational principles of brain processes.

The papers of this special issue reflect the variety in the approaches for the estimation of brain connectivity described above, as well as the need to improve and adapt these approaches to the more and more demanding qualitative requirements and challenges of modern neurophysiological applications. J. Sun et al. offer a review of phase synchronization analysis methods for the inference of functional brain connectivity, describing definitions, estimation, and significance assessment, presenting some extensions, and discussing the issues that affect the detection of phase synchronization from neural data. C. Alvarado-Rojas and M. Le Van Quyen exploit phase synchronization analysis and clustering techniques to identify, from intracranial EEG recordings acquired in epileptic patients during seizure-free periods, dynamic modes of brain synchrony which are characteristic for the wake-sleep cycle. L. Faes et al. introduce a common framework for the unified description of the most popular frequency domain connectivity measures based on linear parametric modeling of multiple time series, discussing their relations, theoretical interpretation, advantages and limitations, and practical estimation. A. Brovelli assesses the reliability of frequency domain Granger causality analysis performed on a single-trial basis, showing that, when combined with parametric statistical tests, Granger causality spectra successfully recover causal interactions in both synthetic and neurophysiological data. Y. Liu and S. Aviyente illustrate the advantages over linear parametric Granger causality of an information theoretic tool, the directed information, as regards the assessment of directional connectivity in both simulated time series and EEG data. D. Marinazzo et al. face the important problem of estimating causality in complex brain networks through fully multivariate approaches; working in the frame of information theory, they provide a novel approach for partial conditioning to a subset of informative variables, showing that this approach can help to overcome computational and numerical problems otherwise arising with traditional full conditioning schemes.

P. Belardinelli et al. investigate in a realistic MEG environment the known localization bias due to correlation between source time series occurring for the popular beamformer inverse method, showing that this bias is relevant only for extremely high degrees of source correlation. F. Avarvand et al. combine subspace and beamformer source localization methods with estimation of the imaginary part of the coherence in simulated and real EEG data, implementing an approach that is sensitive to connectivity rather than to activity and, as such, improves localization accuracy and detection of source interactions. C. Micheli and C. Braun deal with the problem of characterizing connectivity between neural source activity and muscular activity after

the localization of multiple correlated sources; when applied to MEG and electromyographic signals during a pinch grip task, their approach highlights patterns of corticomuscular coherence otherwise not obtainable with the use of standard methods. C. Schmidt et al. propose a new analytical approach to the detection of topological motifs in EEG connectivity networks estimated in patients with major depression during painful stimulation, suggesting that specific motifs may help explaining the relationship between pain and depression. A. Alvarellos-González et al. study structural brain connectivity at the synaptic level, presenting enhanced computational models of neural networks which evidence the potentiation of synaptic connectivity that takes place in the brain when glial cells are considered in addition to artificial neurons.

*Luca Faes
Ralph G. Andrzejak
Mingzhou Ding
Dimitris Kugiumtzis*

Research Article

Fundamental Dynamical Modes Underlying Human Brain Synchronization

Catalina Alvarado-Rojas and Michel Le Van Quyen

Centre de Recherche de l'Institut du Cerveau et de la Moelle Épinière (CRICM), INSERM, UMRS 975 and CNRS UMR 7225, UPMC, Hôpital de la Pitié-Salpêtrière, 75651 Paris cedex 13, France

Correspondence should be addressed to Michel Le Van Quyen, quyen@t-online.de

Received 19 November 2011; Revised 15 January 2012; Accepted 1 March 2012

Academic Editor: Ralph G. Andrzejak

Copyright © 2012 C. Alvarado-Rojas and M. Le Van Quyen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Little is known about the long-term dynamics of widely interacting cortical and subcortical networks during the wake-sleep cycle. Using large-scale intracranial recordings of epileptic patients during seizure-free periods, we investigated local- and long-range synchronization between multiple brain regions over several days. For such high-dimensional data, summary information is required for understanding and modelling the underlying dynamics. Here, we suggest that a compact yet useful representation is given by a state space based on the first principal components. Using this representation, we report, with a remarkable similarity across the patients with different locations of electrode placement, that the seemingly complex patterns of brain synchrony during the wake-sleep cycle can be represented by a small number of characteristic dynamic modes. In this space, transitions between behavioral states occur through specific trajectories from one mode to another. These findings suggest that, at a coarse level of temporal resolution, the different brain states are correlated with several dominant synchrony patterns which are successively activated across wake-sleep states.

1. Introduction

Although much is known about the functional architecture of the brain, its large-scale dynamics remain poorly understood. At this macroscopic level, the existence of large-scale dynamics is confirmed by numerous functional brain mapping studies, showing that multiple distributed cortical areas coordinate their activities during perceptuomotor behavior [1–3]. Phase synchrony is an important candidate for such large-scale integration, mediated by neuronal groups that oscillate in specific bands and enter into precise phase locking over a limited period of time [4, 5]. It is also known that the coordination of these large-scale patterns is subject to drastic changes in different behavioral states, possibly because of state-dependent shifts in both neuromodulatory balance and the thalamic gating of sensory inputs [6–8]. A major challenge is to determine general principles that govern the spontaneous succession of global brain states across the entire wake-sleep cycle. Recently, using simultaneous

local field potentials recorded in multiple forebrain areas in behaving rats, it has been shown that several brain states (e.g., quiet waking, active exploration, and slow-wave sleep) can be represented as distinct clusters in a multidimensional state space representing various levels of local and long distances synchronization [9]. These results suggest that major brain states that comprise the wake-sleep cycle can be identified by a frequency-dependent neuronal cooperativity that involves different oscillatory levels within individual brain regions and transient synchronization across brain areas. Motivated by this study, we investigated the long-term dynamics of human brain synchronization using intracranial recordings of epileptic patients. During the phase of presurgical evaluation, these patients were recorded for up to 7–14 days in order to capture habitual seizures. Invasive EEG recording from intracranial electrodes was required to localize focal epileptic activity and to determine the exact spatial relationships between centers of epileptic activity and functionally significant areas. In contrast to scalp

EEGs, intracranial recordings provide, temporally distant from epileptic seizures, episodes of normal brain activity that are highly differentiated, down to millimeter spatial resolution. The good spatial resolution and the high signal-to-noise ratio offered by intracranial electrodes have been proven valuable in the detection of large-scale dynamical relationships between cortical networks, in both the time and frequency domains and allow a reliable separation of local and long-range mechanisms [5, 10].

2. Methods

2.1. Database. We examined intracranial recordings from 5 subjects with refractory partial epilepsy undergoing presurgical evaluation, hospitalized between February 2002 and July 2007 in the epilepsy unit at the Pitié-Salpêtrière hospital in Paris. Each patient was continuously recorded during several days (duration range, 9–20 days; mean duration, 15 days) with intracranial electrodes (Nicolet acquisition system, CA, USA; 16-bit, bandwidth at 3 dB: 0.1–150 Hz). Signals were digitized at 400 Hz. Depth electrodes were composed of 4 to 10 cylindrical contacts 2.3 mm long, 1 mm in diameter, 10 mm apart center-to-center, mounted on a 1 mm wide flexible plastic probe. Subdural electrodes were strips with 4 to 8 one-sided circular contacts, 2.3 mm in diameter and with a center to center separation of and 10 mm. Pre- and post- implantation MRI scans were evaluated to anatomically and precisely locate each contact along the electrode trajectory. The selection of the sites to implant varied among patients and was made entirely for clinical purposes. Three patients had bilateral depth electrodes placed within the hippocampus in combination with subdural strips added to sample lateral or inferior cortices of the temporal and frontal lobes; two patients had only unilateral depth electrodes in various regions of the neocortex and hippocampus. (Patient 1: 47 channels; unilateral subdural strip electrodes covering the frontal lobe and basal regions of the temporal lobe; unilateral depth electrodes placed in the insula, amygdale, and hippocampus. Patients 2 and 5: 47 and 50 channels; bilateral subdural strip electrodes covering the basal regions of the temporal lobe, lateral, and posterior temporal cortex; bilateral depth electrodes placed in the hippocampus. Patient 3: 32 channels; unilateral depth electrodes placed in the basal regions of the temporal lobe, lateral temporal cortex, and frontal lobe. Patient 4: 38 channels; unilateral depth electrodes placed in basal regions of the temporal lobe, lateral temporal cortex, insula, and amygdala; bilateral depth electrodes placed in the hippocampus). An epileptologist visually evaluated the EEG recordings. Electrodes that exhibited interictal epileptiform discharges (i.e., clearly distinguishable spikes, sharp waves, or spike-and-waves complexes) were identified and were removed from the analysis. Waking and sleep stages were determined by video monitoring and confirmed by the visual inspection of the corresponding EEG recordings. Sleep was here mostly defined as nonrapid eye movement (NREM) periods by the presence of K-complexes (stage 2) or slow waves (stage 3-4).

2.2. Large-Scale Synchronization Analysis. The analysis of phase synchronization between neuronal signals was introduced by [11] to overcome some limitations of conventional methods which cannot disentangle instantaneous amplitudes and phases [10]. The term “synchronization” is used in its strict sense, as a statistical measure of the degree to which two signals are phase locked during a short-time period. Recent studies have demonstrated the ability of this measure to discriminate transient synchronization in intracranial EEG data [12]. Our analysis followed several steps: first, signals from nonoverlapping, consecutive 5-second periods were filtered with a bandpass corresponding to a particular frequency component. Second, the instantaneous phase of each filtered window was extracted by means of the Hilbert transform. Third, the degree of phase locking between a pair of EEG channels was quantified by the trial average of the phase differences on the unit circle in the complex plane:

$$\text{PLV} = \left\| \frac{1}{n} \sum_{t=1}^n e^{i[\phi_1(t) - \phi_2(t)]} \right\|, \quad (1)$$

where n is the number of data points in each time window. This phase-locking value (PLV) varies between 0 (independent signals) and 1 (constant phase lag between the two signals).

Principal Components Analysis. PCA is a method for identifying patterns in data of high dimension and expressing that data in such a way as to highlight their similarities and differences. Since patterns can be hard to find in data of dimension greater than three, where the luxury of graphical representation is not available, PCA is a powerful tool for “visualizing” and “compressing” that data, by reducing the number of dimensions, without much loss of information. PCA usually starts with a large number of data or “observation” vectors, with as many components as system variables. The methods of linear algebra then allow the selection of a special, ordered set of basis vectors, the so-called “principal components.” These vectors are of unit (Euclidean) length and are mutually orthogonal, with pairwise “dot products” equal to zero. The first principal component vector represents a single axis in space. When you project each observation vector onto that axis, the resulting values form a new variable. And the variance of this variable is the maximum among all possible choices of the first axis. The second principal component represents another axis in space, perpendicular to the first. Projecting the observations onto this axis generates another new variable, whose variance is the next largest among all possible choices of this second axis, and so forth. The full set of principal component vectors contains as many elements as the number of original variables. But it is common for the sum of the variances of the first few principal components to exceed, say, 80–90% of the total variance of the original data. By examining plots of these few new variables, researchers often develop a deeper understanding of the driving forces that generated the original data. In mathematical terms, the unordered set of principal component vectors is simply the set of eigenvectors of the covariance matrix of the observation vectors. The first principal component is

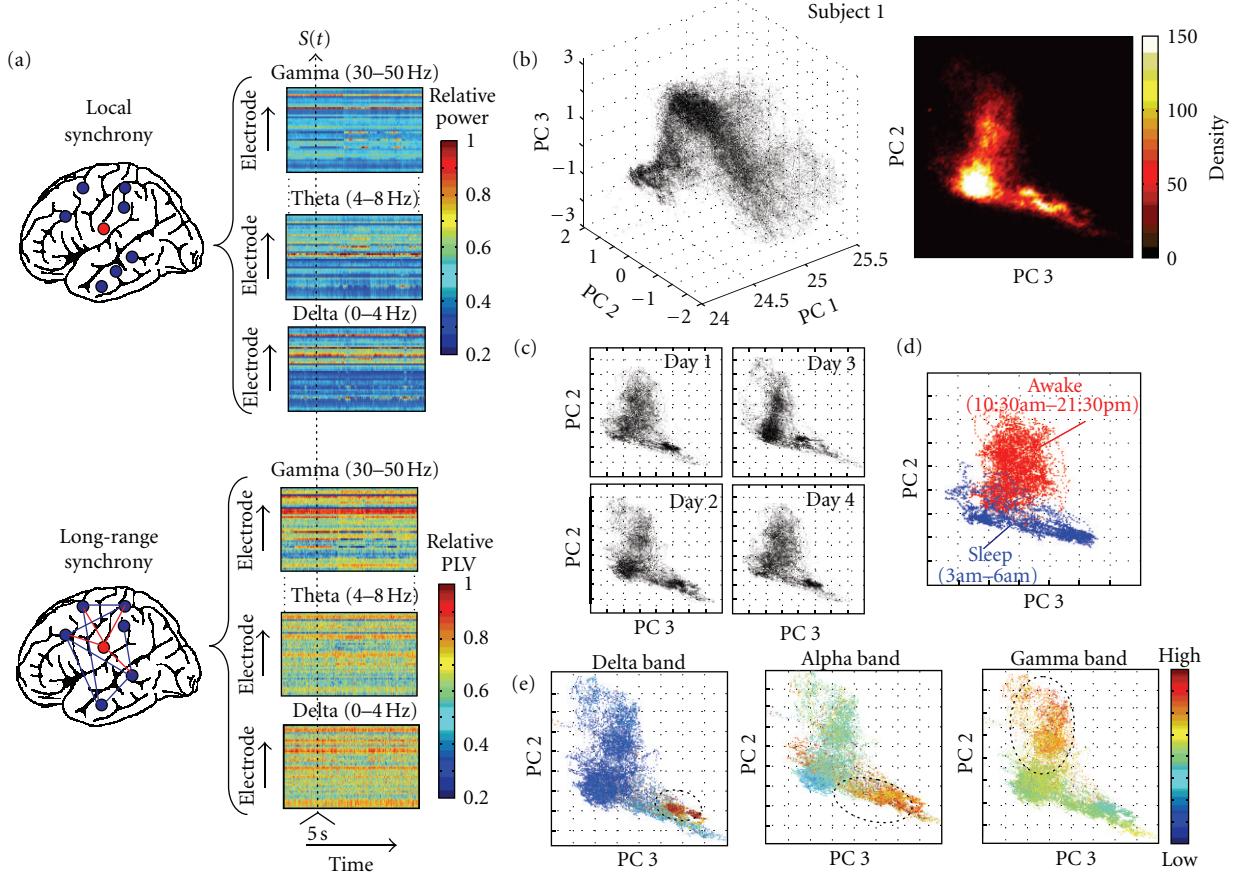


FIGURE 1: After a narrow band filtering of the intracranial EEGs in the delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta 1 (13–20 Hz), beta 2 (20–30 Hz), and gamma (30–50 Hz) frequency bands, local- and long-range synchronizations were, respectively, estimated by the spectral power of each recording contact and by the mean phase-locking values (PLV) between every contact and all the others. This computation allows the characterization of the multifrequency synchronization patterns of each time window t as a vector $S(t)$. (b) Scatter and density plots of 4 successive days (i.e., 96 hours, patient 1), in the space of the first principal components. (c, d) Distributions in the state space across several days and during waking and sleep states. (e) The spectral amplitudes were color coded in the state space, characterizing three main internal frequencies of individual regions in the delta, alpha, and gamma bands.

the eigenvector with the largest eigenvalue, the second principal component corresponds to the next largest eigenvalue, and so on.

Hierarchical Clustering. Hierarchical clustering uses a so-called “agglomerative” approach, in which individual items, perhaps patterns, are joined to form larger groups. These groups are then joined again and again, until the process has been carried to completion, forming a single hierarchical “tree.” This hierarchical clustering proceeds in a simple manner from an initial state, in which each cluster consists of a single item. First, having selected a metric or distance function d , the matrix of distances between all possible pairs (r, s) of clusters ($d(r, s)$) is formed, and the two closest (or “most similar”) clusters are chosen. This is the first true stage in the “clustering” process. If several pairs of clusters have the same separation distance, a predetermined rule is used to decide between alternatives. Second, the two selected clusters are merged to produce a new, larger cluster. Third,

the distances are calculated between this new cluster and all remaining clusters. Fourth, this process continues iteratively, until a single cluster, consisting of all the individual items, remains. Whenever it is necessary to determine the two “closest”, most similar clusters, Ward’s method is used: for every cluster pair (r, s) , the sum of the squares of the distances between all items in their composite cluster and the mean (or centroid) of that cluster is computed [13]. The pair which achieves the minimum of this measure is then selected as the pair to be combined, thus maximizing within-cluster homogeneity. Spurious, small sized clusters were removed. The optimum grouping was defined as the one that minimized the ratio between intracluster and intercluster distance measures, producing the used value for the expected number of clusters. Several other distance measures between clusters (e.g., average, centroid, median distances) were tested with qualitatively similar results. We use the implementations from the Matlab’s Statistics Toolbox (The Mathworks, Natick, MA).

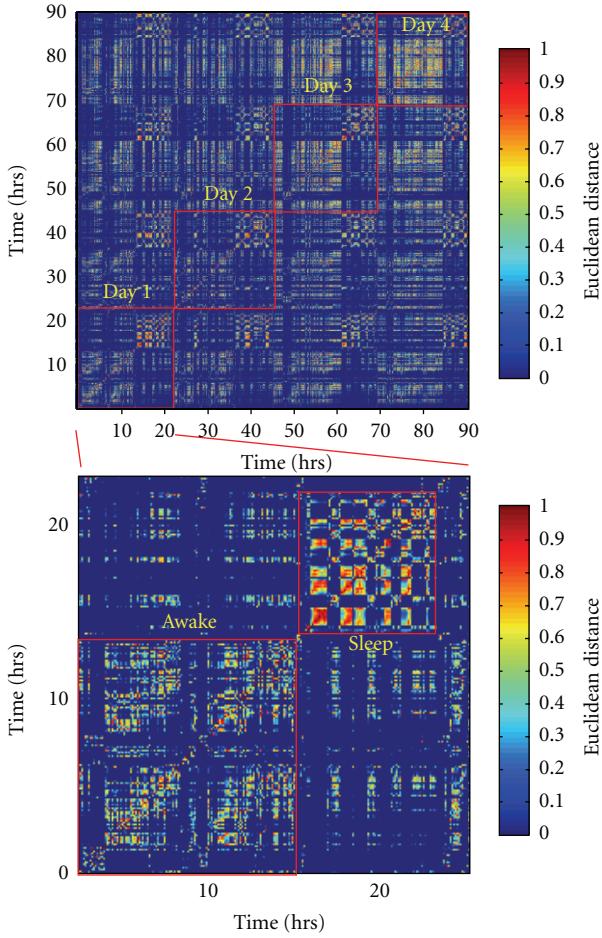


FIGURE 2: Correlation matrix showing the similarity of all the windows compared with each other over 4 successive days (upper map) and during one day (lower map).

State-Space Representation. Following previous work [12], consecutive 5-second EEG recording periods were quantified by a synchronization vector (S) that characterizes the multifrequency synchronization patterns in the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta 1 (13–20 Hz), beta 2 (20–30 Hz), and gamma (30–50 Hz) frequency bands. Higher frequency bands >50 Hz were not considered here because the corresponding activities mainly reflect short-lasting (<100 msec) and low-voltage events, having only a small implication in average synchronizations computed over windows of a few seconds. For every window of time and every frequency band, S quantified two different types of synchronization [5] (Figure 1(a)): (i) local synchronization, estimated by the spectral power of each recording contact and reflecting the frequency-specific summation of coherent currents in a sufficient number of cells that generate externally detectable local field potentials [5] and (ii) long-range synchronization, estimated by the phase-locking values (PLV) for every possible combination of different contacts, characterizing the temporal relationships between different brain regions in a particular frequency band [11]. On the basis of all pairwise computations, we computed for each

contact the average synchronization, defined as the mean PLV between a given contact and all the others. Furthermore, separately for each frequency band, all the values of local and long-range synchronization were normalized by the minimum and maximum over all the recording contacts to produce normalized values bounded between 0 and 1. This normalization factor is used to adjust the data to compensate for experimental variability and to “balance” the values from the local and long-range synchronizations being compared. For 50 channels, the dimension of S is $50 \times 6 \times 2 = 600$, we typically analyzed $\sim 5 \times 10^4$ time windows (2–4 days). To reduce this high-dimensional and possibly redundant data to a lower dimensional space and to reveal hidden underlying factors controlling the dynamics, we used the principal components analysis (PCA). PCA of S led to the determination of the underlying modes that characterize the network states, rank-ordered by their importance, allowing the representation of these states as distinct clusters in a low-dimensional state space [9]. The first three principal components (PCs) of S were used, typically representing around 90% of the total variance. In order to reduce the small-scale variability and to average over microstructures to yield large-scale temporal structures, resulting PCs were further smoothed with a moving average filter of one minute width.

3. Results

We analyzed long-term intracranial recordings of 5 epileptic patients, continuously recorded with a EEG-video monitoring system for successive days (duration ranging from 38 to 103 hours), one day after the implantation of the electrodes and several hours before the first seizure (separated from 10 to 48 hours, depending on the data set). Patients continued taking their standard doses of anticonvulsant medications during this period. The patients had been implanted with both depth electrodes and subdural electrodes distributed over subcortical and neocortical regions. The positioning of electrodes varied among patients (see Section 2.1). We observed that the first three principal components were able to identify and characterize several distinct groups of states across the wake-sleep cycle. While the number of possible synchronization patterns can be very large, we found that most of them occupy an L-shape structure, as can be seen in scatter and density plots (Figure 1(b)). Distributions in the state space were quantitatively similar across several days (Figure 1(c)). Comparative analyses of the behavioral states of the patients showed that the waking and sleep states occupied different regions in the state space (Figure 1(d)). When pooled spectral amplitudes were color coded on this space [9], it was possible to characterize three segregated regions with internal dominant frequencies (Figure 1(e)). The higher gamma power values were observed within a specific region of the state space, mostly associated with the waking state. This observation supports previous reports showing that gamma synchronizations seem to be essential for waking-state information processing. Clearly separated, higher delta powers were exclusively localized in the deep

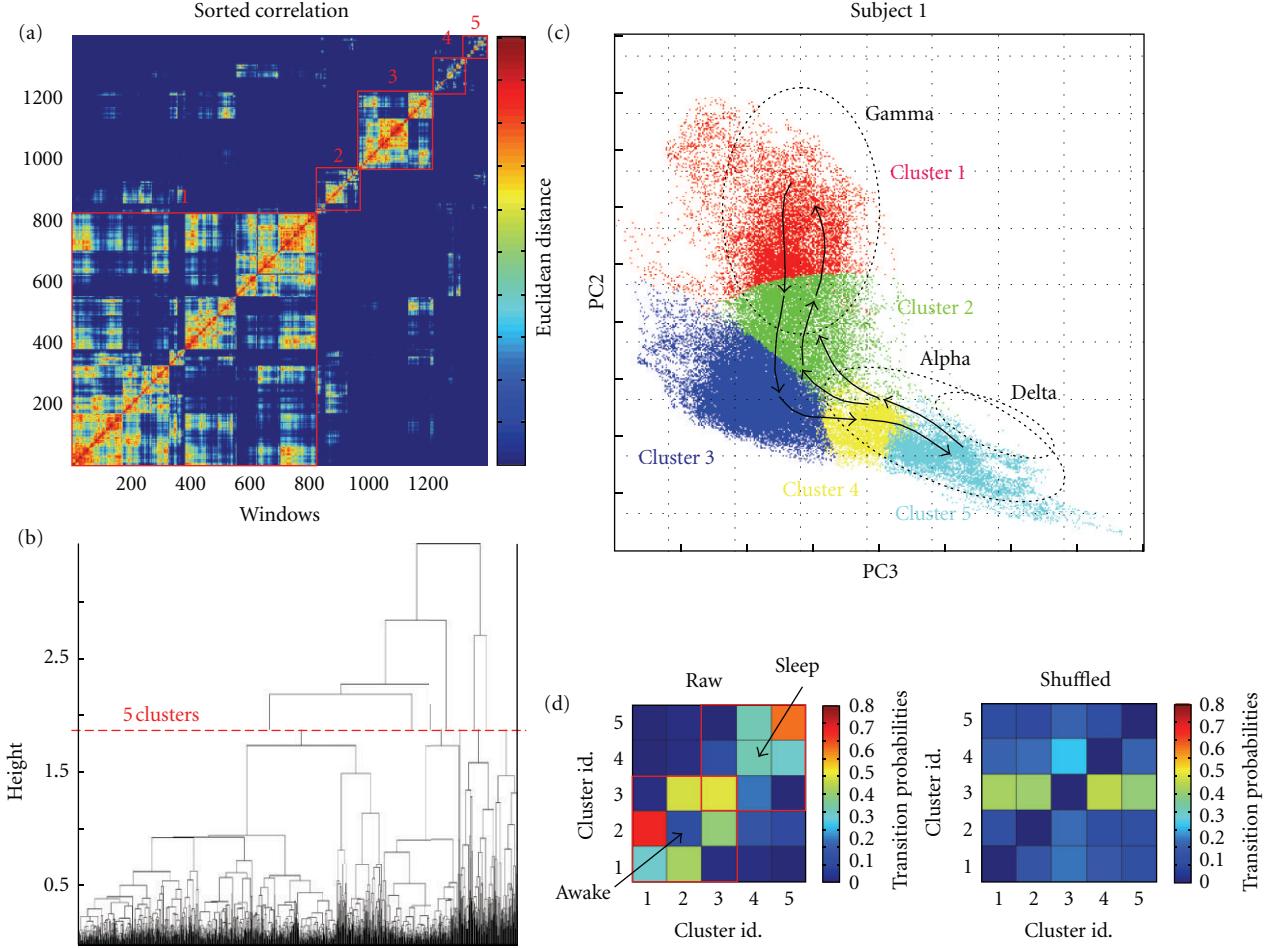


FIGURE 3: (a) Same correlation matrix as in Figure 2, now sorted in order of similarity by the clustering algorithm. (b) Dendrogram of correlation matrix. Levels of the dendrogram correspond to different sets of clusters in the correlation matrix. At the top of the dendrogram, a single branch signifies that all avalanches are in one cluster. Just below this, the dendrogram divides into two branches, representing a set of two clusters. Branching continues further down the dendrogram until every window is in its own family. The red line crosses the dendrogram at minimum of the ratio between intracluster and intercluster distances and indicated that 5 clusters can be identified (note here that one small sized cluster was removed). The 5 corresponding clusters were reported in the sorted correlation matrix (a). (c) The corresponding 5 clusters were reported in the state space, coded by different colors. The probabilities of transition between the different clusters are depicted using different arrow sizes (small: $0.2 < P < 0.4$ and large: $P > 0.4$). Direct cluster-to-cluster transitions were mostly identified between proximal modes in the state space. (c) Matrices of transition probabilities between the characteristic modes defined by clustering, for both the actual and shuffled data.

sleep region, also reflecting a maximal distance to the spectral-coded gamma region. Finally, a distinct frequency region in the range of sleep spindles (12–15 Hz, alpha band in Figure 1(e)) was localized near the region associated with delta/slow oscillations [14]. This remarkable frequency segregation of three domains in the state space confirms that distinct synchronization modes mapped different behavioral states [9]. Similar patterns were not observed using only local synchronizations (i.e., spectral powers), suggesting that long-range synchronizations between distant regions are required.

To investigate these recurrent structures more precisely, all the windows were checked against each other for similarity via the euclidian distance in the PC space (Figure 2). Visual inspection of correlation matrices revealed a large number of positive correlations, suggesting that many of

the windows produced were recurrent over successive days. Hierarchical clustering algorithms were used to rearrange a correlation matrix from temporal order to order of similarity (Figure 3(a)). When presented in order of similarity, several large squares of high-correlation values appeared on the diagonal of the matrix, suggesting that the windows could be grouped into several clusters that were highly similar within themselves. Within several of the large cluster in a sorted matrix, there were smaller subclusters with even greater correlations (Figure 3(a)). The relationships between these clusters and subclusters could be succinctly described by a dendrogram (Figure 3(d)). Each line that cut across the dendrogram at a different level represented a different way of grouping the windows into clusters. The best clustering was defined at the minimum value of the ratio between

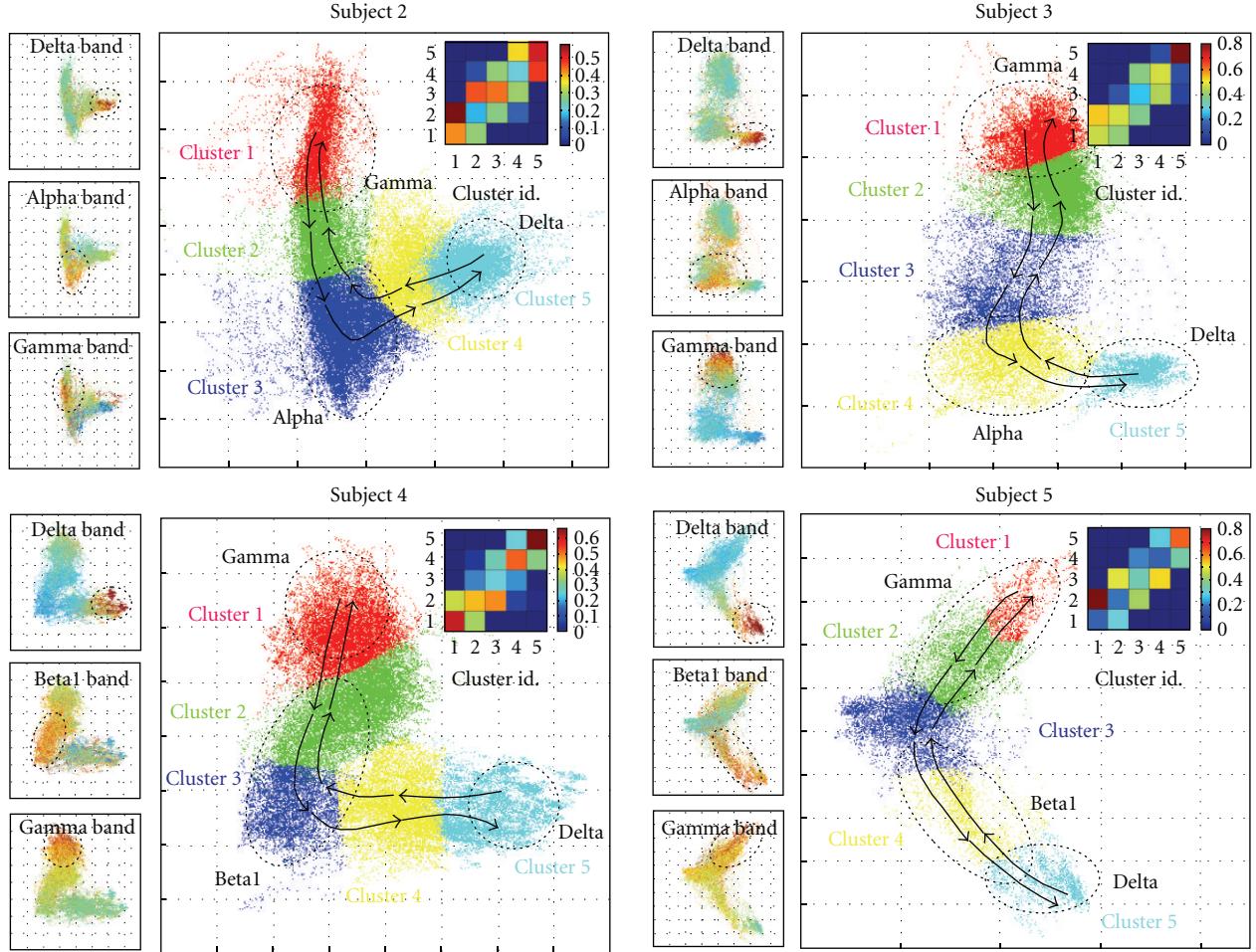


FIGURE 4: Global state space for 4 patients. For each patient, several characteristic clusters were identified by a hierarchical clustering algorithm and were coded by different colors (analyzed periods for patient 2: 103 hours; patient 3: 38.7 hours; patient 4: 86 hours; patient 5: 40 hours). The probabilities of transition between the different clusters are depicted using different arrow sizes (small: $0.2 < P < 0.4$ and large: $P > 0.4$). Note that the global dynamic structures governing the trajectories across the state space are similar among all the different patients (insets: matrices of transition probabilities). In a comparable way, the color-coded spectral state spaces conserved three different segregated regions in the delta, alpha/beta, and gamma bands (first column).

intraccluster and intercluster distances. We found that most synchronization patterns can be reasonably fit into 3–5 characteristic modes (Figures 3(d) and 3(c)).

To further investigate their global dynamic structures, we explored possible causal relationships by deducing the matrix of transition probabilities between the characteristic modes within a time period under 2 minutes (Figure 3(d)). Concerning the inner stability of individual modes, sleep states represent the most stable dynamics, with the strongest probability to remain in the same mode ($P \approx 0.6$). In contrast, we observed a significantly greater inner instability for the modes associated with wakefulness ($P \approx 0.3$). Direct cluster-to-cluster transitions were mostly identified between proximal modes in the state space, those with high probabilities (see the transition matrix in Figure 3(d) and arrows in Figure 3(c) for $P > 0.2$). Direct cluster-to-cluster transitions between nonadjacent modes are rare ($P < 0.1$). Thus the most frequent trajectories are surprisingly

simple, showing, on average, a tendency to follow a flow visiting successively adjacent characteristic modes across wake-sleep states. To measure the statistical significance of these dynamical structures, it is necessary to compare the actual data to what would be caused by chance. We used a shuffling to permute the temporal sequence, preserving the original spatial structures in the PC space, but destroying all dynamical structures. Transition matrices of 50 shuffled data were generated, and a maximum matrix was constructed for each patient. As expected, this transition matrix obtained from the shuffled data hinted that preferred cluster-to-cluster transitions between proximal modes could only be found in the actual data (Figure 3(d)). Surprisingly, although a considerable degree of interpatient variability is to be expected from the different electrode implantations, the L-shaped distributions in the state space were quantitatively similar across the five patients (Figure 4). Furthermore, we found a remarkable similarity in the spectral-coded state space,

including three segregated regions in the delta, alpha/beta, and gamma bands. Finally, the global dynamic structures governing the trajectories between proximal modes are conserved among the different patients.

4. Discussion

Regardless of the relatively small sample size used in the present study, replicability of the results across subjects leads us to believe that global brain states can be mapped into a low-dimensional space based on the degree of local- and long-range synchronization between multiple cortical areas. Automatic cluster analysis made it possible to quantitatively assess the similarities between these different synchronization patterns and to identify functional categories and natural transitions between them. Using this compact representation, we report, with a remarkable similarity across the patients with different locations of electrode placement, that the complex patterns of brain synchrony during the wake-sleep cycle can be represented by a small number of characteristic clusters in which cortical network can dynamically operate. These regimes correspond to distinct global brain states and are correlated with the occurrence of major wake-sleep states.

Although our state-space framework obtained encouraging results, future work should address several issues. In particular, we have mainly studied the ability of a state-space representation in tracking global brain dynamics at low temporal resolution and operating within a few clusters, quantitatively inferred by hierarchical clustering algorithms. One difficulty here is the determination of the minimal number of physiologically meaningful clusters. The problem of dimension reduction is very difficult, especially when the target categories for classification remain unknown. It remains possible that other dimension reduction techniques might provide useful physiological features and identify more dynamic states. Especially, future methods could improve upon the poor temporal resolution in the state-space method because of the smoothing procedure. However, given the relatively slow temporal evolution of behavioral states, such slow temporal dynamics are likely well captured by the state-space framework.

How can this description be useful? The state-space framework proposed here may be helpful for sleep stage scoring (i.e., the process of classifying the different stages of the sleep). Indeed, it is known that most of the current stage-coding approaches, both manual and automatic, face several important limitations [15]. Furthermore, our state-space representation may help to better describe transitions between different sleep stages. Indeed, according to the recommendations of Rechtschaffen and Kales [16], most algorithms to identify wake-sleep states based on EEG features implicitly assume that the wake-sleep cycle consists of several categorically different and stable states. This approach tends to characterize the wake-sleep cycle as a stair case process, jumping back and forth between a set of state. This stair case representation of states promotes the unrealistic view that state transitions occur instantaneously,

with no intermediate periods between them, even when the dynamics of the system do not clearly resemble any predefined states [9]. Finally, in the context of neurology, disturbances of large-scale synchronized networks have been implicated in several brain disorders, such as epilepsy, schizophrenia, autism, and Parkinson's disease [17]. Our state-space representation of global brain dynamics may help to identify pathological alterations in large-scale patterns. In particular, this representation may be helpful in identifying dynamic state fluctuations of the epileptic brain and possibly characterizing long-term pathological transitions to seizures [12, 18].

Overall, our present results strengthen the recent observations in rats that behavioral states and their transitions can be identified by synchronizations across/within forebrain areas [9]. Additionally, our results support the role of oscillation-mediated temporal links, activated differently depending on the ongoing behavioral state, in the coordination of specific information transfer between distant brain regions [19]. Furthermore, following earlier proposals [20, 21], our descriptions make explicit generic structures of large-scale brain dynamics, that is, characteristics that are observed independently of the particular variation of the network under consideration. Together, these results provide new insights into the neurophysiological correlates of state-dependent aspects of human brain synchronization. We anticipate that this type of comprehensive quantification will have powerful applications in the development of automatic recognition of behavioral states [22].

Acknowledgment

This work was partially supported by the European Union-FP7 project EPILEPSIAE (Evolving Platform for Improving Living Expectation of Subjects Suffering from Ictal Events, Grant no 211713).

References

- [1] W. J. Freeman, "Mesoscopic neurodynamics: from neuron to brain," *Journal of Physiology Paris*, vol. 94, no. 5-6, pp. 303–322, 2000.
- [2] A. R. Damasio, "Synchronous activation in multiple cortical regions: a mechanism for recall," *Seminars in The Neuroscience*, vol. 2, pp. 287–296, 1990.
- [3] W. Singer and C. M. Gray, "Visual feature integration and the temporal correlation hypothesis," *Annual Review of Neuroscience*, vol. 18, pp. 555–586, 1995.
- [4] P. R. Roelfsema, A. K. Engel, P. König, and W. Singer, "Visuomotor integration is associated with zero time-lag synchronization among cortical areas," *Nature*, vol. 385, no. 6612, pp. 157–161, 1997.
- [5] F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: phase synchronization and large-scale integration," *Nature Reviews Neuroscience*, vol. 2, no. 4, pp. 229–239, 2001.
- [6] M. Steriade, "Corticothalamic resonance, states of vigilance and mentation," *Neuroscience*, vol. 101, no. 2, pp. 243–276, 2000.

- [7] J. Fell, M. Staedtgen, W. Burr et al., "Rhinal-hippocampal EEG coherence is reduced during human sleep," *European Journal of Neuroscience*, vol. 18, no. 6, pp. 1711–1716, 2003.
- [8] J. L. Cantero, M. Atienza, J. R. Madsen, and R. Stickgold, "Gamma EEG dynamics in neocortex and hippocampus during human wakefulness and sleep," *NeuroImage*, vol. 22, no. 3, pp. 1271–1280, 2004.
- [9] D. Gervasoni, S. C. Lin, S. Ribeiro, E. S. Soares, J. Pantoja, and M. A. Nicolelis, "Global forebrain dynamics predict rat behavioral states and their transitions," *Journal of Neuroscience*, vol. 24, no. 49, pp. 11137–11147, 2004.
- [10] T. H. Bullock, M. C. McClune, J. Z. Achimowicz, V. J. Iragui-Madoz, R. B. Duckrow, and S. S. Spencer, "EEG coherence has structure in the millimeter domain: subdural and hippocampal recordings from epileptic patients," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 3, pp. 161–177, 1995.
- [11] J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase-synchrony in brain signal," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.
- [12] M. Le Van Quyen, J. Soss, V. Navarro et al., "Preictal state identification by synchronization changes in long-term intracranial EEG recordings," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 559–568, 2005.
- [13] J. H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, pp. 236–244, 1963.
- [14] F. Amzica and M. Steriade, "Electrophysiological correlates of sleep delta waves," *Electroencephalography and Clinical Neurophysiology*, vol. 107, no. 2, pp. 69–83, 1998.
- [15] P. S. Jensen, H. Sorensen, H. Leonthin, and P. Jennum, "Automatic sleep scoring in normals and in individuals with neurodegenerative disorders according to new international sleep scoring Criteria," *Journal of Clinical Neurophysiology*, vol. 27, no. 4, pp. 296–302, 2010.
- [16] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*, Public Health Service, U.S. Government Printing Office, Washington, DC, USA, 1968.
- [17] P. J. Uhlhaas and W. Singer, "Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology," *Neuron*, vol. 52, no. 1, pp. 155–168, 2006.
- [18] W. Stacey, M. Le Van Quyen, F. Mormann, and A. Schulze-Bonhage, "What is the present-day EEG evidence for a preictal state?" *Epilepsy Research*, vol. 97, no. 3, pp. 243–251, 2011.
- [19] A. Sirota, J. Csicsvari, D. Buhl, and G. Buzsáki, "Communication between neocortex and hippocampus during sleep in rodents," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 4, pp. 2065–2069, 2003.
- [20] C. A. Skarda and W. J. Freeman, "How brains make chaos in order to make sense of the world," *Behavioral and Brain Sciences*, vol. 10, pp. 161–195, 1987.
- [21] T. M. McKenna, T. A. McMullen, and M. F. Shlesinger, "The brain as a dynamic physical system," *Neuroscience*, vol. 60, no. 3, pp. 587–605, 1994.
- [22] M. K. Kiyimik, M. Akin, and A. Subasi, "Automatic recognition of alertness level by using wavelet transform and artificial neural network," *Journal of Neuroscience Methods*, vol. 139, no. 2, pp. 231–240, 2004.

Research Article

Localizing True Brain Interactions from EEG and MEG Data with Subspace Methods and Modified Beamformers

Forooz Shahbazi Avarvand,^{1,2} Arne Ewald,^{3,4} and Guido Nolte⁵

¹ IDA Group, Fraunhofer Institute FIRST, Kekuléstraße 7, 12489 Berlin, Germany

² Department of Computer Science, Faculty of Mathematics and Natural Sciences II, Humboldt-Universität zu Berlin, Rudower Chaussee 25, 10099 Berlin, Germany

³ Machine Learning Group, Berlin Institute of Technology, Franklinstr 28/29, 10587 Berlin, Germany

⁴ NIRx Medizintechnik GmbH, Baumbachstraße 17, 13189 Berlin, Germany

⁵ Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20246 Hamburg, Germany

Correspondence should be addressed to Forooz Shahbazi Avarvand, forooz.shahbazi@first.fraunhofer.de

Received 21 November 2011; Revised 17 February 2012; Accepted 10 May 2012

Academic Editor: Ralph G. Andrzejak

Copyright © 2012 Forooz Shahbazi Avarvand et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To address the problem of mixing in EEG or MEG connectivity analysis we exploit that noninteracting brain sources do not contribute systematically to the imaginary part of the cross-spectrum. Firstly, we propose to apply the existing subspace method “RAP-MUSIC” to the subspace found from the dominant singular vectors of the imaginary part of the cross-spectrum rather than to the conventionally used covariance matrix. Secondly, to estimate the specific sources interacting with each other, we use a modified LCMV-beamformer approach in which the source direction for each voxel was determined by maximizing the imaginary coherence with respect to a given reference. These two methods are applicable in this form only if the number of interacting sources is even, because odd-dimensional subspaces collapse to even-dimensional ones. Simulations show that (a) RAP-MUSIC based on the imaginary part of the cross-spectrum accurately finds the correct source locations, that (b) conventional RAP-MUSIC fails to do so since it is highly influenced by noninteracting sources, and that (c) the second method correctly identifies those sources which are interacting with the reference. The methods are also applied to real data for a motor paradigm, resulting in the localization of four interacting sources presumably in sensory-motor areas.

1. Introduction

Electroencephalography (EEG) and magnetoencephalography (MEG) are noninvasive measurements of brain activity with an excellent temporal resolution in the order of milliseconds but poor spatial resolution. In the past decades the main focus was the analysis of event-related potentials, that is, the average brain response to a given stimulus. More recently, the variability of brain activity and especially its interpretation as signatures from the brain as a dynamical network has attracted many researchers [1, 2]. A specific expression of variability is the occurrence of neural oscillations which are hypothesized to be a mechanism of functional communication within the brain [3–6]. A

large variety of methods exist to identify interactions of rhythmic activity, including coherence [7], AR modeling [8], Granger causality [9], and methods based on phase couplings [10].

The most serious problem in the interpretation of EEG or MEG in terms of brain interaction arises from the poor spatial resolution. First of all, at the sensor level it is not clear whether a functional relationship between sensors reflects an interaction between two different neural populations or is due to the mixing of sources into sensors. Furthermore, since the inverse problem, that is, the calculation of brain activity from EEG or MEG data, has no unique solution, any estimate is error prone: also estimated source activities are in fact a largely unknown mixture of the true activities. While this

problem, usually termed “artifact of volume conduction,” is well known since a long time [11], it is increasingly addressed lately [12–19]. One major result is that also sophisticated and popular inverse methods like beamformers may produce substantial misinterpretations of the results [18].

To overcome the problem of volume conduction it was suggested to study the imaginary part of coherency [20] because nonvanishing values of that quantity can only be explained by true interactions. Imaginary part of coherency has been used for the estimation of functional connectivity inside the brain in several studies [21–23]. On the sensor level, this allows to establish the presence of brain interactions but only little can be said about the origins of the interactions inside the brain. Analysis on sensor level was further pursued in [24] where a method was proposed to separate pairs of interacting sources from each other. The results are two-dimensional subspaces which contain the topographies, that is the electric potentials or magnetic fields, for each pair of sources. A further decomposition into the topographies of the individual sources is not possible without making spatial assumptions about the sources. In [17] such a decomposition was performed assuming that respective source distributions, estimated as minimum L₂-norm solutions, have minimal spatial overlap. The problem here is that such minimum norm solutions are extremely blurred, and even if the unknown true sources are not overlapping the estimates may have substantial spatial overlap.

In this paper we propose a more natural approach to deal with the case when the outcomes of a sensor level approach are subspaces rather than topographies of individual sources. In fact, subspace methods like the Multiple Signal Classification (MUSIC) and variants of it to be discussed below are designed to find sources which explain subspaces by dipolar sources [25–27]. Such methods are typically applied on low-rank approximations of covariance matrices and work optimally if all sources are independent of each other. To localize interacting sources, we here suggest to simply apply the subspace methods to low-rank approximations of the imaginary part of the cross-spectrum. We show next that this method correctly estimates the source locations also in the presence of strong background noise.

The second question we address is how to estimate with which other source each of the found sources is interacting. This analysis will be based on LCMV beamforming, which is a popular inverse method to analyze EEG or MEG data [12, 28]. Similarly to localization using MUSIC, we will adapt the beamformer to be most sensitive to interactions rather than being sensitive to strong power.

This paper is organized as follows. In Section 2 we will explain the mathematical background on the imaginary part of the cross-spectrum, subspace methods, and on beamformers. In Section 3 we will present our modifications of the RAP-MUSIC approach and of beamformers in order to study interacting brain sources, and in Section 4 we will demonstrate the performance using simulations of two and four interacting sources in the presence of background noise of various strengths as well as real data. A conclusion is given in Section 5.

2. Background

2.1. MUSIC and RAP-MUSIC. Multiple Signal Classification (MUSIC) is a localization method based on dominant subspaces spanned by the vector structure of the data [25]. The general procedure is to divide the vector space of the data into a signal subspace and a noise-only subspace which is orthogonal to the signal subspace. The algorithm is used for acoustic imaging [29, 30] and for the analysis of electrophysiological recordings of brain activity [26, 27]. It finds the source locations as those for which the principle angle between the noise subspace and the forward model of the source is maximum or, equivalently, for which the principle angle between the signal subspace and the forward model is minimal. In a nutshell, the MUSIC algorithm scans all possible source locations and estimates whether a source at each location is consistent with the measured data explicitly including the possibility that several sources are simultaneously active and in general not independent of each other.

We will at first consider the case of fixed dipole orientations. For MUSIC, a subspace of the signal is determined as the space spanned by the set of eigenvectors corresponding to the P largest eigenvalues of the covariance matrix of the data C , which itself has usually full rank for noisy data. P is the (assumed) number of sources, and it is assumed to be substantially smaller than the number of electrodes. (The true number of sources is in general unknown and it is advisable to choose P rather too large than too small.)

We denote the forward model of the dipole at location q_i as $L_{M \times 1}$, where M is the number of electrodes. In order to estimate the consistency between the forward model of a given grid point and the subspace, the angle between them is calculated as

$$\cos^2 \theta(L, \phi) = \frac{L^T \phi \phi^T L}{L^T L}, \quad (1)$$

where $(\cdot)^T$ denotes transpose, and ϕ is the matrix of the P largest eigenvalues of the covariance matrix of the data. We note that formulations using angles between model and noise subspace are formally equivalent. Formulations using signal subspace only are computationally more efficient since the dimensionality of the signal subspace is lower compared to the dimensionality of the noise subspace.

The angle θ is calculated in all the grid points and the forward model corresponding to the minimum angle is estimated as the dipole pattern. If, as in EEG or MEG, for each grid point several forward solutions exist, corresponding to three different dipole orientations, the source orientation corresponding to the largest value of $\cos \theta$ is chosen.

The main disadvantage of MUSIC is that finding several maxima is difficult when the number of sources increases [27]. As a remedy, several modifications of MUSIC are proposed which are based on the idea of localizing the sources sequentially [27, 31–33]. One of the variants proposed by Mosher and Leahy [27] is a modification called Recursively Applied and Projected (RAP)-MUSIC. Here, instead of searching simultaneously for several local maxima, only global maxima are determined iteratively. In order to find the next source location, the subspace is updated by

projecting out the previously found topographies and then the maximization is repeated.

To be explicit, let L_k for $k = 1 \dots n - 1$ be the set of patterns of the $n - 1$ previously found sources. In order to find the location of the n th source, the new subspace is defined by removing the patterns both from the forward models and the subspace estimation. The projection matrix for the n th source estimation reads

$$P = I - A(A^T A)^{-1} A^T, \quad (2)$$

where $A = [L_1 L_2 \dots L_{n-1}]$ is the matrix containing as columns all the previously found dipole patterns.

Similarly to the first MUSIC scan, the angle between the forward model at each grid point and the subspace is calculated while the forward models and the subspace are updated by projecting out the previous source patterns. Therefore

$$\cos^2 \theta(L_P, \phi_P) = \frac{L_P^T \phi_P \phi_P^T L_P}{L_P^T L_P}, \quad (3)$$

where $L_P = PL$ and $\Phi_P = \text{ortho}(P\Phi)$ where $\text{ortho}(E)$ orthonormalizes the columns of a matrix E .

The algorithm performs as many iterations as the predefined number of sources.

For unknown dipole orientations an optimization over orientation is included in the calculation of the angle θ . Then

$$L = \hat{L}\alpha, \quad (4)$$

where \hat{L} is an $N \times 3$ matrix containing as columns the topographies of dipoles in x -, y - and z -direction and α is a 3×1 vector. Then the angle is given as

$$\cos^2 \theta(L_P, \phi_P) = \max_{\alpha} \frac{\alpha^T \hat{L}_P^T \phi_P \phi_P^T \hat{L}_P \alpha}{\alpha^T \hat{L}_P^T \hat{L}_P \alpha}, \quad (5)$$

with $\hat{L}_P = P\hat{L}$. The maximization can be done analytically, and α is given by the eigenvector corresponding to the maximum eigenvalue of

$$D \equiv (\hat{L}_P^T \hat{L}_P)^{-1} \hat{L}_P^T \phi_P \phi_P^T \hat{L}_P. \quad (6)$$

Note that the RAP-MUSIC search results both in location and orientation of the sources.

2.2. Imaginary Part of Cross-Spectrum. A covariance matrix is a measure of linear coupling between two signals in the time domain. The analogue in the Fourier domain is the complex valued cross-spectrum which reflects the linear coupling of the signals for all frequencies. Due to the artifacts of volume conduction, it is not always easy to differentiate between the real connectivities and the ones caused by volume conduction. Nolte et al. [20] suggest that the imaginary part of coherency, which is in fact the normalized imaginary part of cross-spectrum, is a measure robust to artifacts of volume conduction in the sense that a nonvanishing imaginary part cannot be explained by independent sources regardless

of the number of sources and how they are mapped into sensors provided that this mapping is essentially instantaneous which is in fact an excellent approximation for frequencies below 1 KHz [34].

Coherency between two EEG channels i and j is defined as

$$C_{ij}(f) = \frac{S_{ij}(f)}{(S_{ii}(f)S_{jj}(f))^{1/2}}, \quad (7)$$

where $S_{ij}(f)$ is the cross-spectrum of the two channels at frequency f and is defined as

$$S_{ij}(f) = \langle x_i(f)x_j^*(f) \rangle, \quad (8)$$

where $x_i(f)$ and $x_j(f)$ are the (complex) Fourier transformations of the time series of $x_i(t)$ and $x_j(t)$ of channels i and j , respectively, $\langle \cdot \rangle$ is the expectation value, and $*$ is complex conjugation. In practice, the expectation value is obtained by averaging over a large number of epochs. $S_{ii}(f)$ and $S_{jj}(f)$ are the autospectra of the signals at channels i and j , respectively.

2.3. Beamformers. The goal of beamforming is to estimate the time course $s(t)$ of a dipole at a specific location in the brain as accurate as possible. To achieve that goal, sensor data are linearly combined such that the (presumed) activity of other sources is minimized [35]. We here recall shortly the basic procedure.

If the location where we want to calculate the time course is q_i and the activity of the dipole at this location is $s_i(t)$, the data $\mathbf{X}(t)$ measured with EEG electrodes is the superposition of N dipoles at sampling time t :

$$\mathbf{X}(t) = \sum_{i=1}^N \mathbf{g}_i s_i(t) + n(t), \quad (9)$$

where \mathbf{g}_i is the forward model for a source at location q_i with given orientation. The vector \mathbf{g}_i is of size $M \times 1$, where M is the number of electrodes and $n(t)$ is additive noise which is assumed here to arise from noninteracting sources. A beamformer is a spatial filter constructed to (a) pass the signal from the source of interest with unit magnitude and (b) to minimize total power. If the source of interest is independent of all other sources (called background), power values of source of interest and background are additive. Therefore, minimizing total power is equivalent to minimizing the power of the background and hence to maximizing signal-to-noise ratio. At a specific location inside the brain we estimate the signals $y_j(t)$, for $j = 1, 2, 3$, corresponding to the source component in x -, y -, and z -direction, as

$$y_j(t) = \mathbf{W}_j^T \mathbf{X}(t). \quad (10)$$

If we denote the topography of the source in direction j as \mathbf{g}_j the filter weights W_j are chosen to satisfy the following constraint:

$$\min_{\mathbf{W}_j} (\mathbf{W}_j^T \mathbf{C} \mathbf{W}_j) \quad (11)$$

$$\text{subject to } \mathbf{W}_j^T \mathbf{g}_j = 1, \quad (12)$$

where the matrix C is the covariance matrix of the measured data in time domain or the cross-spectrum matrix in frequency domain. The optimization is solved by

$$\mathbf{W}_j = C^{-1} \mathbf{g}_j \left[\mathbf{g}_j^T C^{-1} \mathbf{g}_j \right]^{-1}. \quad (13)$$

So far, all dipole components were calculated, and such a beamformer is called “vector beamformer.” To specify the direction at each grid point, it is a common approach to maximize the power. We only mention this without going into details, because we are interested in observing interacting and not necessarily strong sources. The respective choice of orientation will be explained in Section 3.

One point which should be considered in beamforming is the correlation between the sources. Due to the presence of correlated sources, the estimated variance of the source of interest is significantly less than the true value. Therefore, a modified version of LCMV called Nulling Beamformer [12, 36] was suggested forcing an additional nulling constraint in order to make sure that the influence of the sources at specific other locations and orientations is suppressed. We recall the procedure for a set of sources with given orientation. Combining the nulling constraint with the unit gain condition (12) in LCMV results in

$$\mathbf{W}_i^T G = \mathbf{f}_i^T, \quad (14)$$

where $G = [\mathbf{g}_1, \dots, \mathbf{g}_N]$ contains as columns the topographies of N sources and

$$\mathbf{f}_i = [0 \dots 0 1 0 \dots 0]^T, \quad (15)$$

is a vector whose i th element is one and the rest are zero. Solving the equation, using Lagrange multipliers, results in

$$\mathbf{W}_i = C^{-1} G \left[G^T C^{-1} G \right]^{-1} \mathbf{f}_i. \quad (16)$$

The obtained nulling beamformer gain has a unit gain at the location of interest, zero gains at a small set of given locations other than the location of interest and minimizes the power for the i th source.

We finally note that a vector beamformer is often formulated as a nulling beamformer for which for each dipole location and direction j both other orthogonal directions were nulled out.

An LCMV beamformer is really a two-step procedure with two different rationals. In the first step, spatial filters are designed to estimate brain activity for each location in the brain as clean as possible. This step is not a localization approach. The localization is done in the second step by defining the most interesting sources as those which have strongest power. Below, we will use only the first step of the beamformer formulation because we are interested only in interacting sources which are not necessarily the ones with strongest power.

3. New Methods

3.1. Getting Subspaces from Imaginary Part of Cross-Spectrum (CS). The standard way to define the subspace of the

data used for the MUSIC algorithm, as we discussed in Section 2.1, is to calculate the eigenvectors of the covariance matrix of the data. We suggest to replace the covariance matrix by the imaginary part of the cross-spectrum of the data at a specific frequency in RAP-MUSIC. As we discussed in Section 2.2, the imaginary part of the cross-spectrum is inconsistent with noninteracting sources. Since we are interested in localizing the interacting sources, we defined the subspace of the data based on the imaginary part of the cross-spectrum just to make sure that noninteracting sources like noise do not appear in localization results.

3.2. Maximizing Imaginary Coherence in Subspaces. According to the definition, coherency between two EEG channels i and j is equivalent to the complex valued cross-spectrum normalized by the power in the channels. In order to calculate the coherency between the source location i and any other grid point in the brain, the signals originating from these locations will be calculated using a beamformer. The moment of a dipole at location j is therefore calculated as

$$y(f) = \mathbf{AX}(f), \quad (17)$$

where $\mathbf{X}(f)$ is the Fourier transform of the data at frequency f and filter weights, A , are calculated using either LCMV beamformer or nulling beamformer. Let $G = [\mathbf{g}_1, \dots, \mathbf{g}_N]$ be the matrix of dipole patterns of the dipoles estimated at the source locations by RAP-MUSIC then the activity of the source, $s(f)$, in the frequency domain at the i th location is

$$s(f) = \mathbf{V}_i^T \mathbf{X}(f), \quad \mathbf{V}_i = (C_R)^{-1} G \left[G^T (C_R)^{-1} G \right]^{-1} \mathbf{f}_i. \quad (18)$$

The filter weights at each source location \mathbf{V}_i are calculated using nulling beamformer weights in (16) and the vector \mathbf{f}_i is defined in (16). The filter weights could also be estimated using LCMV approach in (13) but in order to reduce the interaction of other sources in the estimation of the time course of the source of interest using nulling beamformer is preferred. The matrix C in (16) and (13) is replaced by the real part of cross-spectrum matrix of the data, C_R , in order to have real-valued weights instead of complex ones.

In a similar approach, the activity of each grid point is calculated using (17) where the filter weights, A , are calculated based on LCMV beamforming weights in (13) in directions x , y , and z . In the classical LCMV beamformer, the direction of the dipole is chosen as the direction which maximizes the power of the signal at the corresponding location but in a new approach, we suggest to choose the direction of the dipole not based on the maximum power but on the maximum imaginary part of coherency between a reference and the dipole of interest.

Assuming $z(f)$ is the moment of the dipole at frequency f at location j in the direction of maximum coherency with location i (the seed location) then,

$$z(f) = \alpha^T \mathbf{AX}(f), \quad (19)$$

where α is the vector of size 3×1 which gives us the (yet unknown) direction in which the imaginary part of coherency is maximum.

The cross-spectrum between the estimated source at location i , $s(f)$, and the activity in direction of maximum coherency at location j , $z(f)$, is defined as

$$\langle s(f)z^*(f) \rangle = \mathbf{V}_i^T \langle \mathbf{X}(f)\mathbf{X}^H(f) \rangle A^T \alpha, \quad (20)$$

where $\langle \cdot \rangle$ denotes the expectation value and $*$ is the complex conjugation. The autospectrum of the signal s at location i is defined as

$$\langle s(f)s^*(f) \rangle = \mathbf{V}_i^T \langle \mathbf{X}(f)\mathbf{X}^H(f) \rangle \mathbf{V}_i, \quad (21)$$

where $\langle \mathbf{X}(f)\mathbf{X}^H(f) \rangle$ is equal to the cross-spectrum of the data. Similarly, the cross-spectrum of $z(f)$ at location j is defined as

$$\langle z(f)z^*(f) \rangle = \alpha^T A \langle \mathbf{X}(f)\mathbf{X}^H(f) \rangle A^T \alpha. \quad (22)$$

The imaginary part of coherency then reads

$$\begin{aligned} f(\alpha) &= \frac{\langle s(f)z^*(f) \rangle}{\langle s(f)s^*(f) \rangle \langle z(f)z^*(f) \rangle} \\ &= \frac{\mathbf{V}_i^T C_I A^T \alpha}{(\mathbf{V}_i^T C \mathbf{V}_i)^{1/2} (\alpha^T A C A^T \alpha)^{1/2}} \end{aligned} \quad (23)$$

where C_I denotes the imaginary part of cross-spectrum of $\mathbf{X}(f)$. Let us rename $(\mathbf{V}_i^T C \mathbf{V}_i)$ as D_1 , $(\alpha^T A C A^T \alpha)$ as D_2 , and $\mathbf{V}_i^T C_I A^T \alpha$ as N . In order to maximize $f(\alpha)$, we set the derivative of $f(\alpha)$ to zero:

$$\frac{\partial f}{\partial \alpha} = \frac{(\mathbf{V}_i^T C_I A^T)^T}{D_1^{1/2} D_2^{1/2}} - \frac{1}{2} \frac{N}{D_1^{1/2} D_2^{1/2}} \frac{\partial}{\partial \alpha} (\alpha^T A C_R A^T \alpha) = 0, \quad (24)$$

where C_R is the real part of C . Solving the equation results in

$$\alpha = (A C_R A^T)^{-1} A C_I \mathbf{V}_i. \quad (25)$$

Substituting α in (23) gives us the maximum of the imaginary part of coherency at frequency f . We used the above maximization of coherency after applying the RAP-MUSIC algorithm to the data in order to study the interactions between the localized sources.

4. Results

4.1. Simulations. In this Section, we present the simulations in which we compared the RAP-MUSIC results in the case the subspace of the data is defined with the largest eigenvalues of real part of the cross-spectrum to the case that the subspace is defined based on the imaginary part of the cross-spectrum. We also demonstrate the results of finding the interaction between the sources after being localized by RAP-MUSIC.

In the first simulation, two interacting sources are produced at 10 Hz with the sampling frequency of 100 Hz. The interaction was simulated simply as a delay: if $x_1(t)$ is

the signal of the first source, white noise narrowband filtered at 10 Hz, then the signal of the second source reads $x_2(t) = x_1(t - \tau)$ with the delay set as $\tau = 20$ ms. The total length of the data is 300 sec and is divided into segments of 100 samples each. Each segment has 50 samples overlapping with the previous segment. Additional noise resembling the real brain noise is added to the data. This is done by simulating independent white noise source distributed evenly across the entire brain. The noise was scaled such that the power at 10 Hz at the strongest signal channel was equal for noise and signal of interest.

For the forward solution, calculated using expansions of the electric lead field [37], we used a realistic head model based on a segmented head model taken from the program CURRY. In this simulation, the dimension of the subspace, that is, P , was chosen to be equal to two.

Figures 1 and 2 are illustrations of the estimated source locations using the real and imaginary parts of cross-spectrum. For each voxel, the result of $1/(1 - |\cos \theta|)$ is color coded. The blue circles in the figures represent the true locations of the simulated sources. Figure 1 shows the two estimated sources using imaginary part of the cross-spectrum. In the first RAP-MUSIC step, both of the two sources are localized simultaneously. In the next step the first source is projected out and only one of the sources has remained. Comparing the results in Figure 1 and RAP-MUSIC based on the real part of cross-spectrum in Figure 2 shows that the localization accuracy increases massively when the imaginary part is used to estimate the sources. In fact, we can see source location estimations in Figure 2 which do not fit our true locations. These estimations are the locations where noninteracting sources, noise in this case, are dominant.

Figure 3 shows the estimation error for two and four interacting sources, which are located at random positions inside the head and had random orientations, for 200 independent simulations. We did the simulations with three different noise levels: (a) no noise, (b) low noise corresponding to equal power of noise and signal of interest at 10 Hz averaged over all channels, and (c) high noise corresponding to equal power of noise and signal of interest at 10 Hz at the channel with largest power of the signal of interest.

On the x -axis the localization error of RAP-MUSIC based on the imaginary part of cross-spectrum is shown and the y -axis represents the estimation error resulting from the real part of cross-spectrum. To identify estimated source locations with true ones we calculated the mean distance across all permutations and chose the one which minimized this mean. The error localizations based on the real part of the cross-spectrum are considerably larger than for the imaginary part of cross-spectrum. In fact, by considering the imaginary part of cross-spectrum, we reduce the effect of the noninteracting noise sources on the localization of the interacting sources.

In order to study the connectivity of the sources, we proceeded the simulations by applying the nulling beamformer to the EEG data and maximizing the imaginary part of coherency between the estimated source locations obtained from RAP-MUSIC and all other grid points. We demonstrate

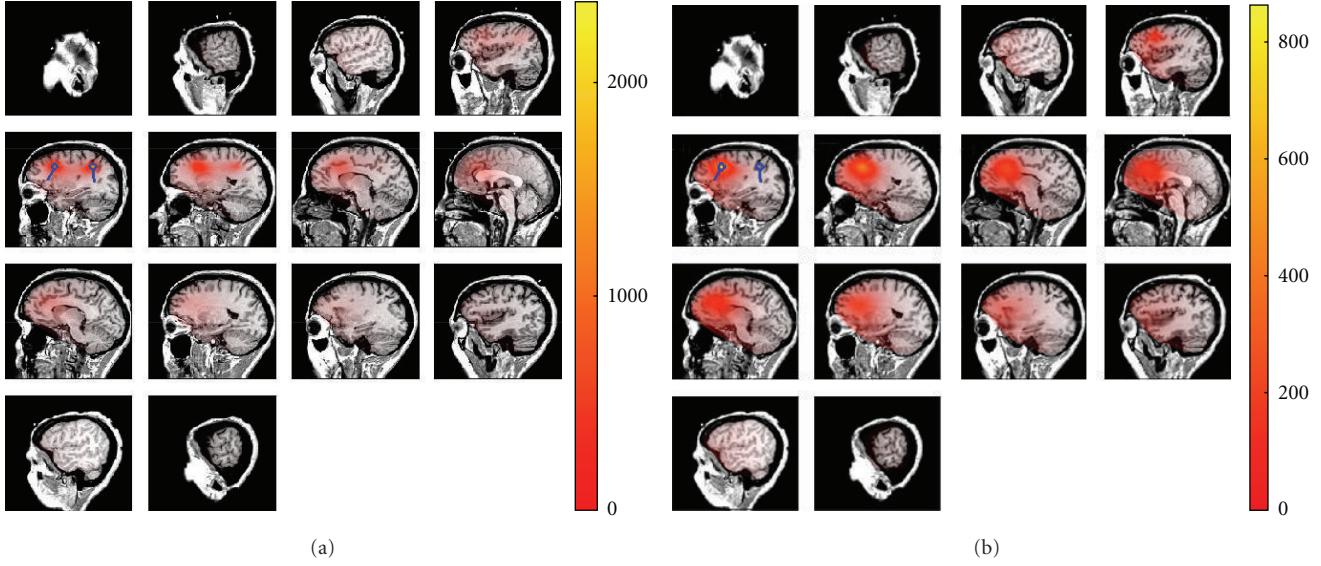


FIGURE 1: Estimated locations of the sources using RAP-MUSIC based on the imaginary part of cross-spectrum. True locations of the sources are shown as blue dipoles and the estimated locations are shown using the heat map. (a) represents the sources after applying the first iteration of RAP-MUSIC and (b) shows the source location after projecting out the global maximum of (1) in the previous iteration.

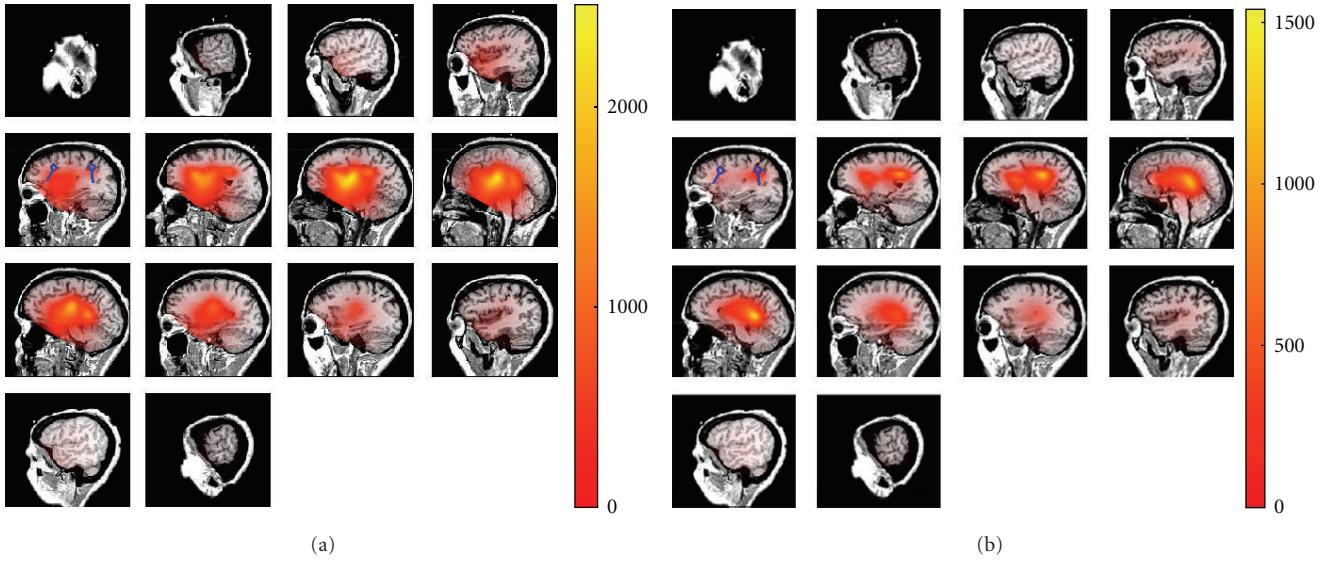


FIGURE 2: Estimated locations of the sources using RAP-MUSIC based on the real part of cross-spectrum. True locations of the sources are shown as blue dipoles and the estimated locations are shown using the heat map. (a) represents the sources after applying the first iteration of RAP-MUSIC and (b) shows the source location after projecting out the global maximum of (1) in the previous iteration.

a typical outcome for a case consisting of four dipoles, two on the left and two on the right hemisphere with interactions within but not across hemispheres and a high noise level (Figure 4). For illustrative purposes, all dipoles were chosen to be in one axial plane, and, although the reconstruction was done in the entire brain, we show only this plane. The results fulfilled our expectations in the way that the highest imaginary part of coherency occurred almost at the same position as the true interacting source positions.

4.2. Real Data. We applied RAP-MUSIC to the real data measured during the imagined hand movement [38] in order to localize four interacting sources as well as their interactions. The cross-spectrum has a dominating alpha rhythm at 10 Hz which is not induced by the task but is considered to be an ongoing activity present at the eyes-open condition as well. The data contains central alpha also at 10 Hz due to event-related synchronization which in this case is induced by the absence of the foot movement which has been the

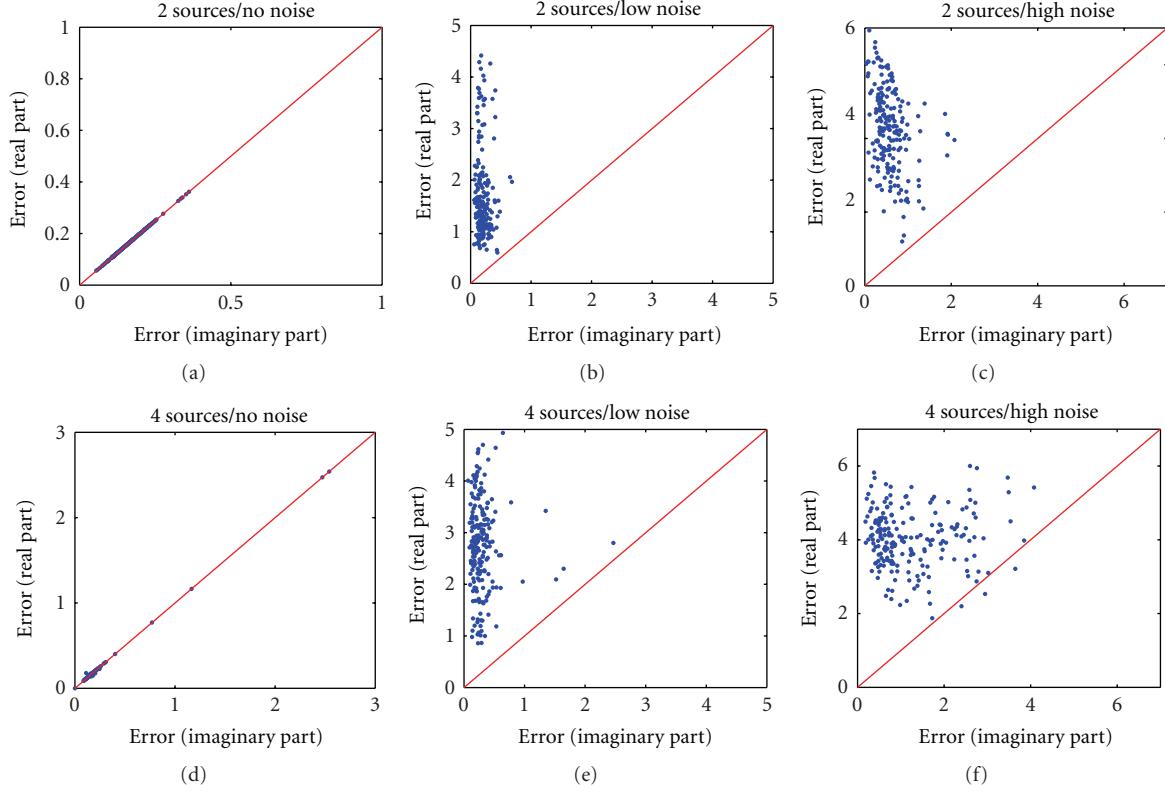


FIGURE 3: Estimation error in RAP-MUSIC based on the real and imaginary part of the cross-spectrum. (a), (b), and (c) show the simulations of two interacting sources in 200 independent simulation cases in random locations. (d), (e), and (f) represent four sources containing two couples of interacting sources simulated in 200 independent simulation cases in random locations. The simulations were done for three different noise levels as explained in the text and as indicated in the figure.

task in nonanalyzed trials. The number of EEG channels is equal to 118 and the number of trials to 70 each with the duration of 3.5 s. The cross-spectrum at 10 Hz is measured with the frequency resolution of 2 Hz. Channel locations were matched on a realistic standard head model taken from the program CURRY (Neuroscan, Hamburg, Germany).

In order to apply RAP-MUSIC based on the imaginary part of cross-spectrum we set the dimension of the subspace equal to four. The source localization resulting in Figure 5 shows two sources in occipital lobe representing the alpha rhythm. Two other sources are more close to the motor cortex which represent the absence of imagined foot movement. To study the interaction between the sources using beamformers and maximization of imaginary coherency described in Section 3, the imaginary part of coherency is maximized between each reference location found in RAP-MUSIC and each grid point. According to the results in Figure 6, the dipoles in each lobe are interacting locally with each other.

5. Conclusion

We adapted two well-established methods, the RAP-MUSIC approach and the LCMV beamformer approach, to localize and characterize interacting brain sources from rhythmic

EEG or MEG data. To study brain interactions robust to artifacts of volume conduction, it is convenient to analyze the imaginary part of the cross-spectrum which is unbiased by noninteracting sources. In contrast to covariance matrices or complex cross-spectra the imaginary part, being antisymmetric, is necessarily degenerate: all singular values occur in pairs, and, for example, a singular value decomposition is not capable to extract the topographies of the individual sources in sensor space even if the true topographies are orthogonal. This is a principle limitation when analyzing interacting sources where dynamical assumptions like statistical independence, as is done for ICA, are inconsistent with the object which is studied. Rather than individual topographies results are naturally subspaces, for which subspace methods are ideal candidates to find the respective sources. In simulations we have shown that RAP-MUSIC, applied on subspaces given by imaginary parts of cross-spectra, properly recovers source locations also in the presence of strong correlated background noise, which was assumed to be generated by noninteracting sources.

This was shown for two and four sources, but not for three. The case of having an odd number of sources differs substantially from the case of an even number. The rank of an antisymmetric matrix is always even and we can only observe in the data an unknown two-dimensional projection of the three-dimensional subspace spanned by all three

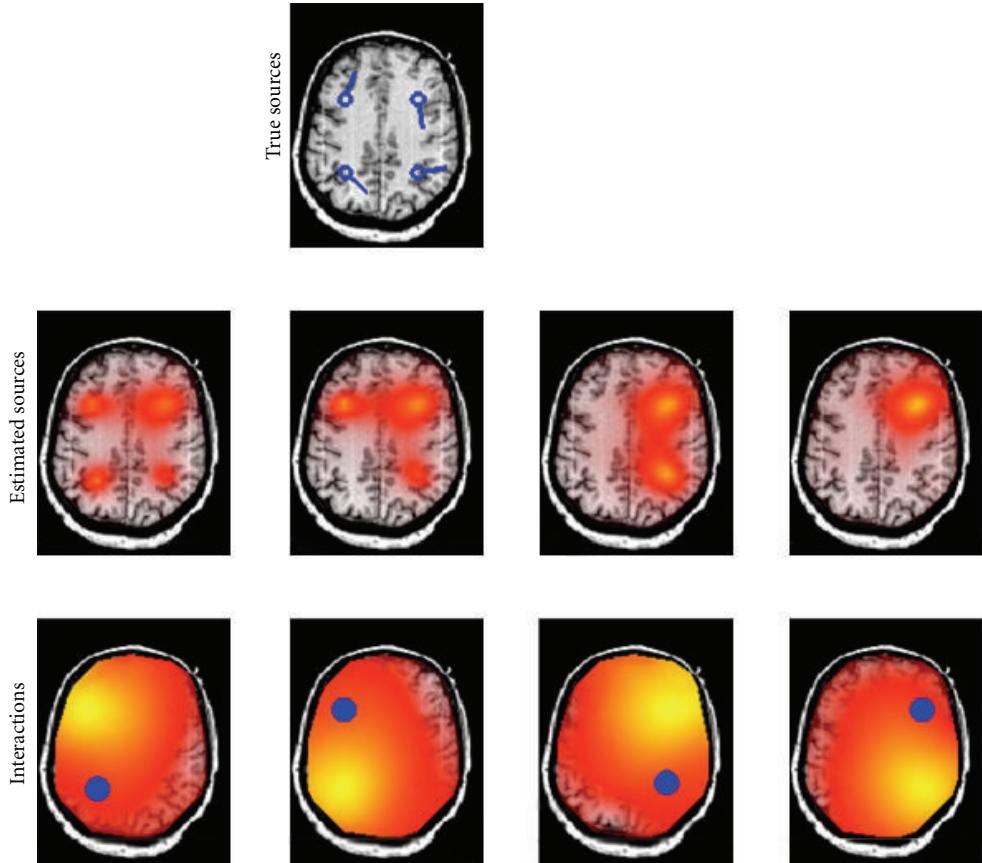


FIGURE 4: Four interacting dipoles are simulated. Each two sources in one hemisphere are interacting with each other. In the top figure, the location of true sources are shown. In the middle row, estimated source locations using RAP-MUSIC based on the imaginary part of cross-spectrum are demonstrated and in the bottom row the area which is interacting with each of the corresponding sources (visualized as blue dots) is demonstrated.

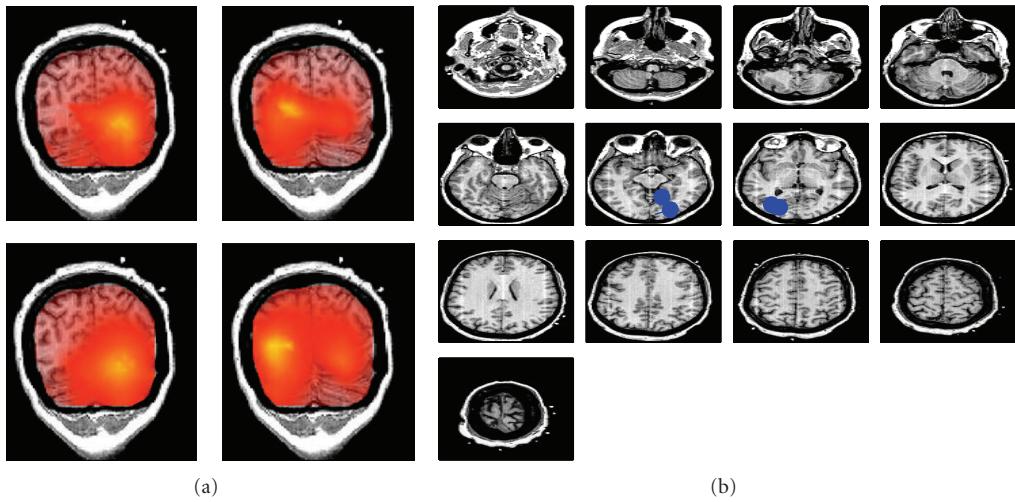


FIGURE 5: (a) RAP-MUSIC applied to the real data to localize 4 interacting sources. Each panel in (a) represents an iteration in RAP-MUSIC starting from the top left panel, then top right and so forth (b) The blue dipoles represent the final estimated source locations after all iterations.

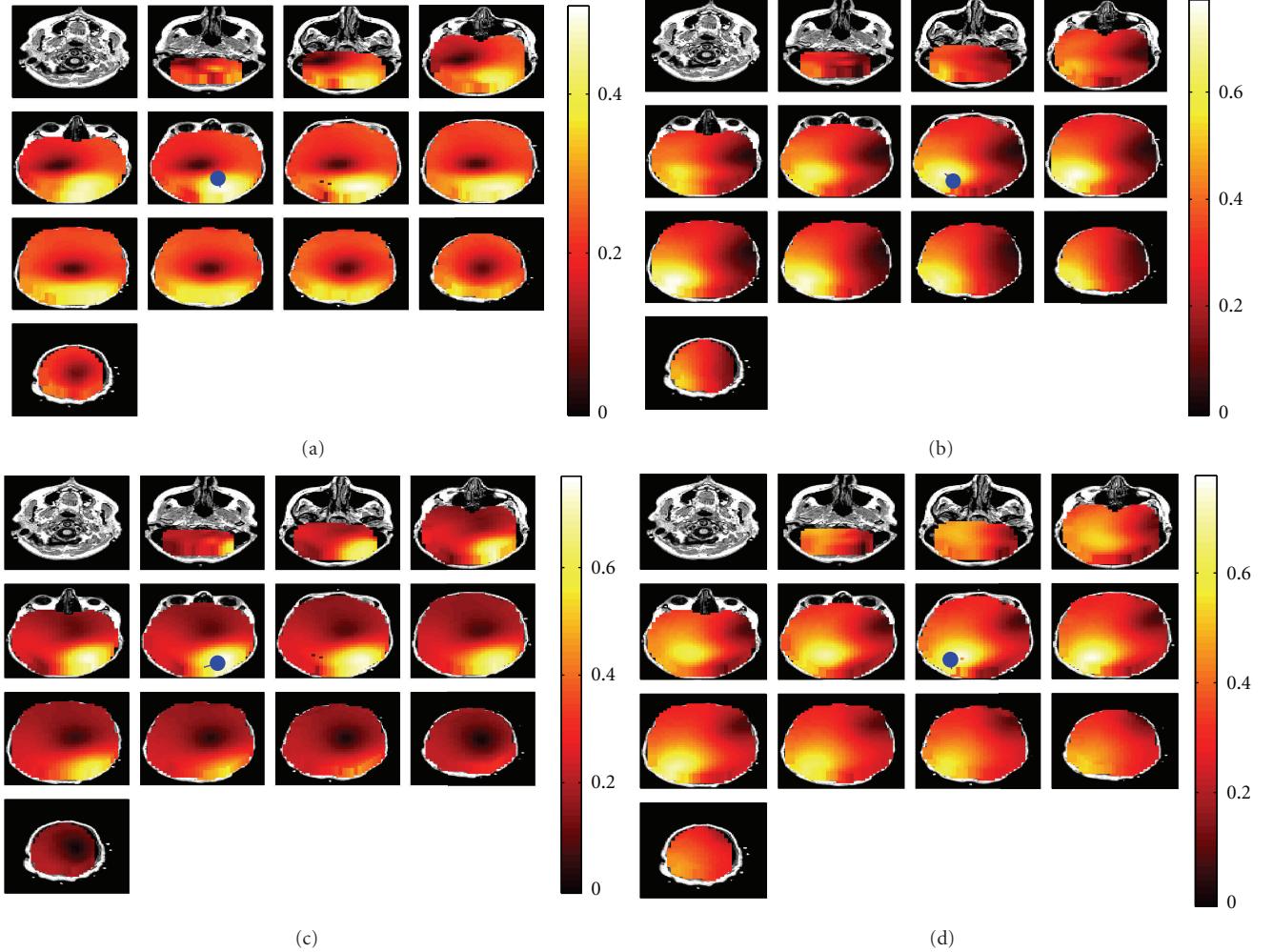


FIGURE 6: Localization of the interaction between four sources using the maximization of imaginary part of coherency. The blue dipoles represent the estimated location of each source which is taken as the reference and the heat map represents the area that is interacting with the corresponding reference location.

topographies. The presented RAP-MUSIC approach is in general not capable to localize sources properly in that case. This problem will be addressed in future work.

To estimate the interaction pattern we adapted the well-known LCMV beamformer to our needs. In “classical” beamformer algorithms the orientation for a given dipole is chosen as the one which maximizes the power in that brain voxel such that the solution picks the strongest source. This was replaced by choosing the direction to maximize the imaginary part of coherency between that voxel and a given reference. To avoid confounding effects by assigning interactions to wrong voxels we also chose to use the Nulling beamformer which sets additional constraints to explicitly exclude contributions from a given set of topographies. This set was defined in terms of the pair of voxels of which the interaction is calculated but a generalization to include other sources is straight forward.

An important advantage of studying the imaginary part of cross-spectra to localize interacting brain sources is that it is applicable without any modification also to differences

of cross-spectra estimated, for example, in two different measurement conditions. An analogous property for the characterization of the interaction, that is, the question which source is interacting with which other, is not possible within the proposed scheme because coherence loses its meaning and is eventually even ill-defined when cross-spectra are normalized with power differences rather than powers. How to characterize interaction from difference of cross-spectra only will be addressed in future work.

Acknowledgments

This project was funded by the Fraunhofer Society of Germany and by the Deutsche Forschungsgemeinschaft (DFG: SFB 936).

References

- [1] K. J. Miller, K. E. Weaver, and J. G. Ojemann, “Direct electrophysiological measurement of human default network areas,”

- Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 29, pp. 12174–12177, 2009.
- [2] P. Fries, “Neuronal gamma-band synchronization as a fundamental process in cortical computation,” *Annual Review of Neuroscience*, vol. 32, pp. 209–224, 2009.
 - [3] P. Fries, “A mechanism for cognitive dynamics: neuronal communication through neuronal coherence,” *Trends in Cognitive Sciences*, vol. 9, no. 10, pp. 474–480, 2005.
 - [4] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie, “The brainweb: phase synchronization and large-scale integration,” *Nature Reviews Neuroscience*, vol. 2, no. 4, pp. 229–239, 2001.
 - [5] A. K. Engel, P. Fries, and W. Singer, “Dynamic predictions: oscillations and synchrony in top-down processing,” *Nature Reviews Neuroscience*, vol. 2, no. 10, pp. 704–716, 2001.
 - [6] W. Singer, “Neuronal synchrony: a versatile code for the definition of relations?” *Neuron*, vol. 24, no. 1, pp. 49–65, 1999.
 - [7] M. J. Brookes, J. R. Hale, J. M. Zumer et al., “Measuring functional connectivity using MEG: methodology and comparison with fcMRI,” *NeuroImage*, vol. 56, no. 3, pp. 1082–1104, 2011.
 - [8] L. Astolfi, F. Cincotti, D. Mattia et al., “Comparison of different cortical connectivity estimators for high-resolution EEG recordings,” *Human Brain Mapping*, vol. 28, no. 2, pp. 143–157, 2007.
 - [9] S. L. Bressler and A. K. Seth, “Wiener-Granger causality: a well established methodology,” *NeuroImage*, vol. 58, no. 2, pp. 323–329, 2011.
 - [10] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, “Measuring phase synchrony in brain signals,” *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.
 - [11] P. L. Nunez, R. B. Silberstein, Z. Shi et al., “EEG coherency II: experimental comparisons of multiple measures,” *Clinical Neurophysiology*, vol. 110, no. 3, pp. 469–486, 1999.
 - [12] H. B. Hui, D. Pantazis, S. L. Bressler, and R. M. Leahy, “Identifying true cortical interactions in MEG using the nulling beamformer,” *NeuroImage*, vol. 49, no. 4, pp. 3161–3174, 2010.
 - [13] J. J. González, S. Mañas, L. de Vera et al., “Assessment of electroencephalographic functional connectivity in term and preterm neonates,” *Clinical Neurophysiology*, vol. 122, no. 4, pp. 696–702, 2011.
 - [14] G. Gómez-Herrero, M. Atienza, K. Egiazarian, and J. L. Cantero, “Measuring directional coupling between EEG sources,” *NeuroImage*, vol. 43, no. 3, pp. 497–508, 2008.
 - [15] C. J. Stam, G. Nolte, and A. Daffertshofer, “Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources,” *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007.
 - [16] G. Nolte, A. Ziehe, V. V. Nikulin et al., “Robustly estimating the flow direction of information in complex physical systems,” *Physical Review Letters*, vol. 100, no. 23, Article ID 234101, 4 pages, 2008.
 - [17] 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, 2008.
 - [18] J. M. Schouffelen and J. Gross, “Source connectivity analysis with MEG and EEG,” *Human Brain Mapping*, vol. 30, no. 6, pp. 1857–1865, 2009.
 - [19] K. Sekihara, J. P. Owen, S. Trisno, and S. S. Nagarajan, “Removal of spurious coherence in MEG source-space coherence analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 11, Article ID 5976430, pp. 3121–3129, 2011.
 - [20] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, “Identifying true brain interaction from EEG data using the imaginary part of coherency,” *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.
 - [21] A. G. Guggisberg, S. M. Honma, S. S. Dalal et al., “Mapping functional connectivity in patients with brain lesions,” *Annals of Neurology*, vol. 63, no. 2, pp. 193–203, 2008.
 - [22] J. Martino, S. M. Honma, A. M. Findlay et al., “Resting functional connectivity in patients with brain tumors in eloquent areas,” *Annals of Neurology*, vol. 69, no. 3, pp. 521–532, 2011.
 - [23] A. Ewald, L. Marzetti, F. Zappasodi, F. C. Meinecke, and G. Nolte, “Estimating true brain connectivity from EEG/MEG data invariant to linear and static transformations in sensor space,” *NeuroImage*, vol. 60, no. 1, pp. 476–488, 2011.
 - [24] G. Nolte, F. C. Meinecke, A. Ziehe, and K. R. Müller, “Identifying interactions in mixed and noisy complex systems,” *Physical Review E*, vol. 73, no. 5, Article ID 051913, 2006.
 - [25] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
 - [26] J. C. Mosher, P. S. Lewis, and R. M. Leahy, “Multiple dipole modeling and localization from spatio-temporal MEG data,” *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 6, pp. 541–557, 1992.
 - [27] J. C. Mosher and R. M. Leahy, “Source localization using recursively applied and projected (RAP) MUSIC,” *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 332–340, 1999.
 - [28] K. Sekihara, S. S. Nagarajan, D. Poeppel, A. Marantz, and Y. Miyashita, “Application of an MEG eigenspace beamformer to reconstructing spatio-temporal activities of neural sources,” *Human Brain Mapping*, vol. 15, no. 4, pp. 199–215, 2002.
 - [29] H. Wang and M. Kaveh, “Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 823–831, 1985.
 - [30] H. W. Chen and J. W. Zhao, “Coherent signal-subspace processing of acoustic vector sensor array for DOA estimation of wideband sources,” *Signal Processing*, vol. 85, no. 4, pp. 837–847, 2005.
 - [31] S. K. Oh and C. K. Un, “Improved MUSIC algorithm for high-resolution array processing,” *Electronics Letters*, vol. 25, no. 22, pp. 1523–1525, 1989.
 - [32] P. Handel, P. Stoica, and A. Nehorai, “Improved sequential MUSIC,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 4, pp. 1230–1239, 1995.
 - [33] G. Bouleux and R. Boyer, “Zero-forcing based sequential MUSIC algorithm,” in *Proceedings of 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '07)*, pp. III1017–III1020, April 2007.
 - [34] J. G. Stinstra and M. J. Peters, “The volume conductor may act as a temporal filter on the ECG and EEG,” *Medical and Biological Engineering and Computing*, vol. 36, no. 6, pp. 711–716, 1998.
 - [35] B. D. van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki, “Localization of brain electrical activity via linearly constrained minimum variance spatial filtering,” *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 867–880, 1997.
 - [36] S. S. Dalal, K. Sekihara, and S. S. Nagarajan, “Modified beamformers for coherent source region suppression,” *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1357–1363, 2006.
 - [37] G. Nolte and G. Dassios, “Analytic expansion of the EEG lead field for realistic volume conductors,” *Physics in Medicine and Biology*, vol. 50, no. 16, pp. 3807–3823, 2005.

- [38] B. Blankertz, G. Dornhege, C. Schäfer et al., “Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 127–131, 2003.

Research Article

Causal Information Approach to Partial Conditioning in Multivariate Data Sets

D. Marinazzo,¹ M. Pellicoro,^{2,3,4} and S. Stramaglia^{2,3,4}

¹ Department of Data Analysis, Faculty of Psychology and Pedagogical Sciences, University of Gent, 9000 Gent, Belgium

² Dipartimento Interateneo di Fisica “Michelangelo Merlin”, University of Bari, 70126 Bari, Italy

³ TIRES-Center of Innovative Technologies for Signal Detection and Processing, University of Bari, 70125 Bari, Italy

⁴ INFN, Sezione di Bari, 70125 Bari, Italy

Correspondence should be addressed to D. Marinazzo, daniele.marinazzo@ugent.be

Received 2 November 2011; Revised 15 March 2012; Accepted 18 March 2012

Academic Editor: Dimitris Kugiumtzis

Copyright © 2012 D. Marinazzo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When evaluating causal influence from one time series to another in a multivariate data set it is necessary to take into account the conditioning effect of the other variables. In the presence of many variables and possibly of a reduced number of samples, full conditioning can lead to computational and numerical problems. In this paper, we address the problem of partial conditioning to a limited subset of variables, in the framework of information theory. The proposed approach is tested on simulated data sets and on an example of intracranial EEG recording from an epileptic subject. We show that, in many instances, conditioning on a small number of variables, chosen as the most informative ones for the driver node, leads to results very close to those obtained with a fully multivariate analysis and even better in the presence of a small number of samples. This is particularly relevant when the pattern of causalities is sparse.

1. Introduction

Determining how the brain is connected is a crucial point in neuroscience. To gain better understanding of which neurophysiological processes are linked to which brain mechanisms, structural connectivity in the brain can be complemented by the investigation of statistical dependencies between distant brain regions (functional connectivity) or of models aimed to elucidate drive-response relationships (effective connectivity). Advances in imaging techniques guarantee an immediate improvement in our knowledge of structural connectivity. A constant computational and modelling effort has to be done in order to optimize and adapt functional and effective connectivity to the qualitative and quantitative changes in data and physiological applications. The paths of information flow throughout the brain can shed light on its functionality in health and pathology. Every time that we record brain activity we can imagine that we are monitoring the activity at the nodes of a network. This activity is dynamical and sometimes chaotic. Dynamical networks [1] model physical and biological behaviour in many

applications; also, synchronization in dynamical network is influenced by the topology of the network itself [2]. A great need exists for the development of effective methods of inferring network structure from time series data; a dynamic version of the Bayesian networks has been proposed in [3]. A method for detecting the topology of dynamical networks, based on chaotic synchronization, has been proposed in [4].

Granger causality has become the method of choice to determine whether and how two time series exert causal influences on each other [5, 6]. This approach is based on prediction: if the prediction error of the first time series is reduced by including measurements from the second one in the linear regression model, then the second time series is said to have a causal influence on the first one. This frame has been used in many fields of science, including neural systems [7–10], reochaos [11], and cardiovascular variability [12].

From the beginning [13, 14], it has been known that if two signals are influenced by third one that is not included in the regressions, this leads to spurious causalities, so an extension to the multivariate case is in order. The conditional Granger causality analysis (CGCA) [15] is based on

a straightforward expansion of the autoregressive model to a general multivariate case including all measured variables. CGCA has been proposed to correctly estimate coupling in multivariate data sets [16–19]. Sometimes though, a fully multivariate approach can entrain problems that can be purely computational or even conceptual: in the presence of redundant variables the application of the standard analysis leads to underestimation of causalities [20].

Several approaches have been proposed in order to reduce dimensionality in multivariate sets, relying on generalized variance [16], principal components analysis [19], or the Granger causality itself [21].

In this paper we will address the problem of partial conditioning to a limited subset of variables, in the framework of information theory. Intuitively, one may expect that conditioning on a small number of variables should be sufficient to remove indirect interactions if the connectivity pattern is sparse. We will show that this subgroup of variables might be chosen as the most informative for the driver variable and describe the application to simulated examples and a real data set.

2. Materials and Methods

We start by describing the connection between the Granger causality and information-theoretic approaches like the transfer entropy in [22]. Let $\{\xi_n\}_{n=1,\dots,N+m}$ be a time series that may be approximated by a stationary Markov process of order m , that is, $p(\xi_n \mid \xi_{n-1}, \dots, \xi_{n-m}) = p(\xi_n \mid \xi_{n-1}, \dots, \xi_{n-m-1})$. We will use the shorthand notation $X_i = (\xi_i, \dots, \xi_{i+m-1})^\top$ and $x_i = \xi_{i+m}$, for $i = 1, \dots, N$, and treat these quantities as N realizations of the stochastic variables X and x . The minimizer of the risk functional

$$R[f] = \int dX dx (x - f(X))^2 p(X, x) \quad (1)$$

represents the best estimate of x , given X , and corresponds [23] to the regression function $f^*(X) = \int dx p(x \mid X)x$. Now, let $\{\eta_n\}_{n=1,\dots,N+m}$ be another time series of simultaneously acquired quantities, and denote $Y_i = (\eta_i, \dots, \eta_{i+m-1})^\top$. The best estimate of x , given X and Y , is now $g^*(X, Y) = \int dx p(x \mid X, Y)x$. If the generalized Markov property holds, that is,

$$p(x \mid X, Y) = p(x \mid X), \quad (2)$$

then $f^*(X) = g^*(X, Y)$ and the knowledge of Y does not improve the prediction of x . Transfer entropy [22] is a measure of the violation of 2: it follows that the Granger causality implies nonzero transfer entropy [24]. Under the Gaussian assumption it can be shown that the Granger causality and transfer entropy are entirely equivalent and just differ for a factor two [25]. The generalization of the Granger causality to a multivariate fashion, described in the following, allows the analysis of dynamical networks [26] and to discern between direct and indirect interactions.

Let us consider n time series $\{x_\alpha(t)\}_{\alpha=1,\dots,n}$; the state vectors are denoted:

$$X_\alpha(t) = (x_\alpha(t-m), \dots, x_\alpha(t-1)), \quad (3)$$

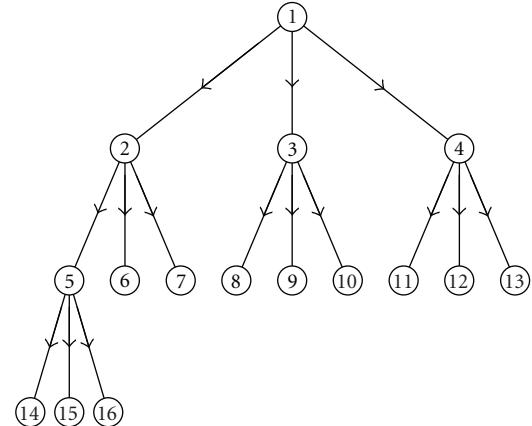


FIGURE 1: A directed rooted tree of 16 nodes.

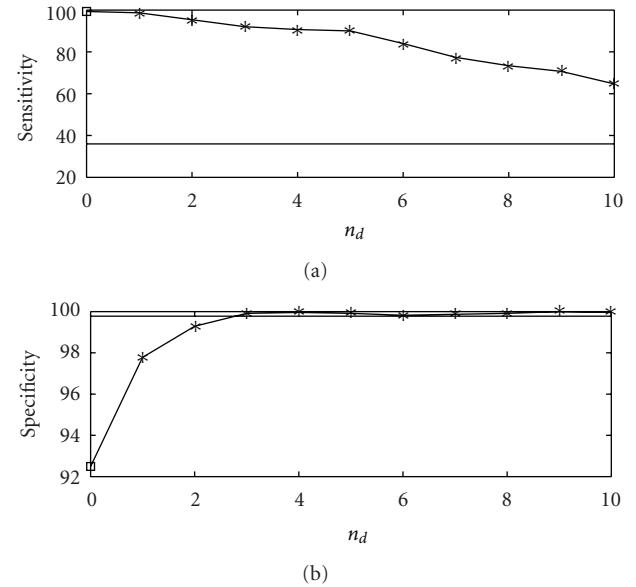


FIGURE 2: The sensitivity (a) and the specificity (b) are plotted versus n_d , the number of variables selected for conditioning, for the first example, the rooted tree. The number of samples N is 100, and the order is $m = 2$; similar results are obtained varying m . The results are averaged over 100 realizations of the linear dynamical system described in the text. The empty square, in correspondence to $n_d = 0$, is the result from the bivariate analysis. The horizontal line is the outcome from multivariate analysis, where all variables are used for conditioning.

m being the window length (the choice of m can be done using the standard cross-validation scheme). Let $\epsilon(x_\alpha \mid \mathbf{X})$ be the mean squared error prediction of x_α on the basis of all the vectors \mathbf{X} (corresponding to linear regression or nonlinear regression by the kernel approach described in [24]). The multivariate Granger causality index $c(\beta \rightarrow \alpha)$ is defined as follows: consider the prediction of x_α on the basis of all the variables but X_β and the prediction of x_α using all the

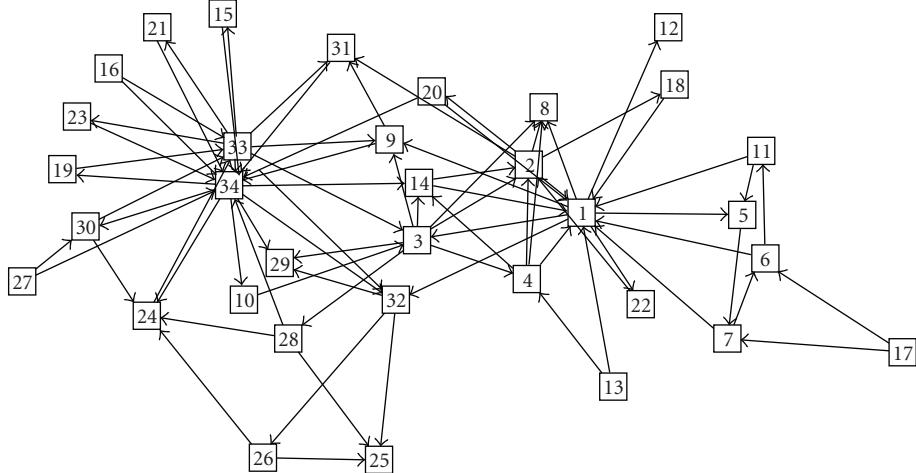
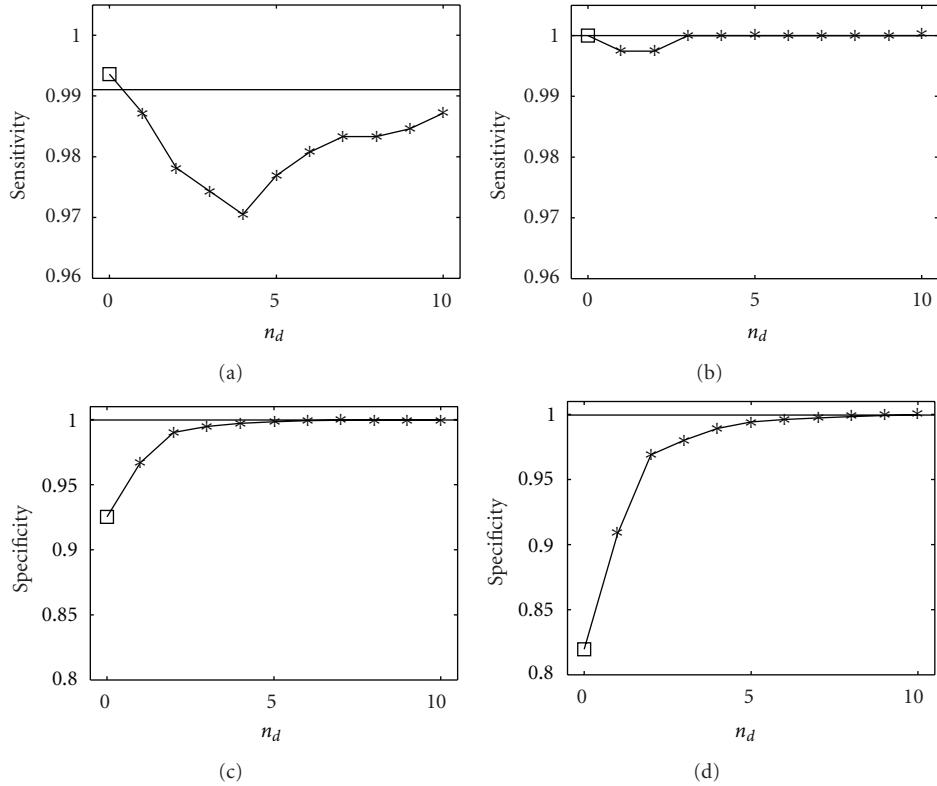


FIGURE 3: The directed network of 34 nodes obtained assigning randomly a direction to links of the Zachary network.

FIGURE 4: Sensitivity and specificity are plotted versus n_d , the number of variables selected for conditioning, for two values of the number of samples N , 500 (left) and 1000 (right). The order is $m = 2$, similar results are obtained varying m . The results are averaged over 100 realizations of the linear dynamical system described in the text. The empty square, in correspondence to $n_d = 0$, is the result from the bivariate analysis. The horizontal line is the outcome from multivariate analysis, where all variables are used for conditioning.

variables, then the causality measures the variation of the error in the two conditions, that is,

$$c(\beta \rightarrow \alpha) = \log \frac{\epsilon(x_\alpha | \mathbf{X} \setminus X_\beta)}{\epsilon(x_\alpha | \mathbf{X})}. \quad (4)$$

Note that in [24] a different definition of causality has been used,

$$\delta(\beta \rightarrow \alpha) = \frac{\epsilon(x_\alpha | \mathbf{X} \setminus X_\beta) - \epsilon(x_\alpha | \mathbf{X})}{\epsilon(x_\alpha | \mathbf{X} \setminus X_\beta)}. \quad (5)$$

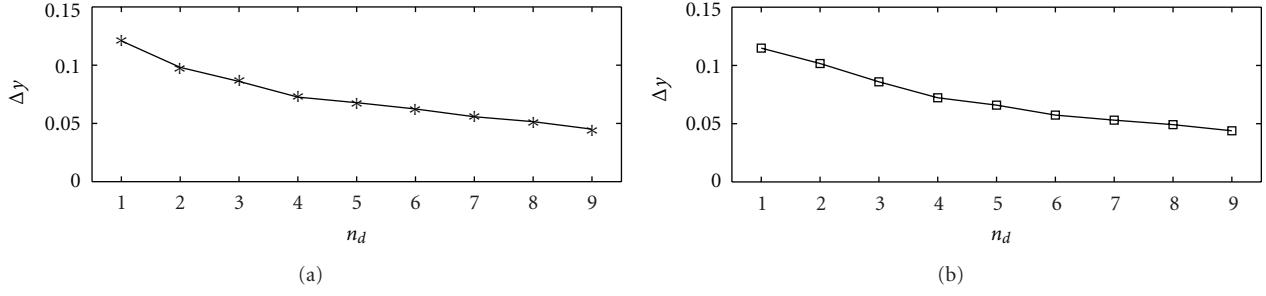


FIGURE 5: The mutual information gain, when the $(n_d + 1)$ th variable is included, is plotted versus n_d for two values of the number of samples N , 500 (a) and 1000 (b). The order is $m = 2$. The information gain is averaged over all the variables.

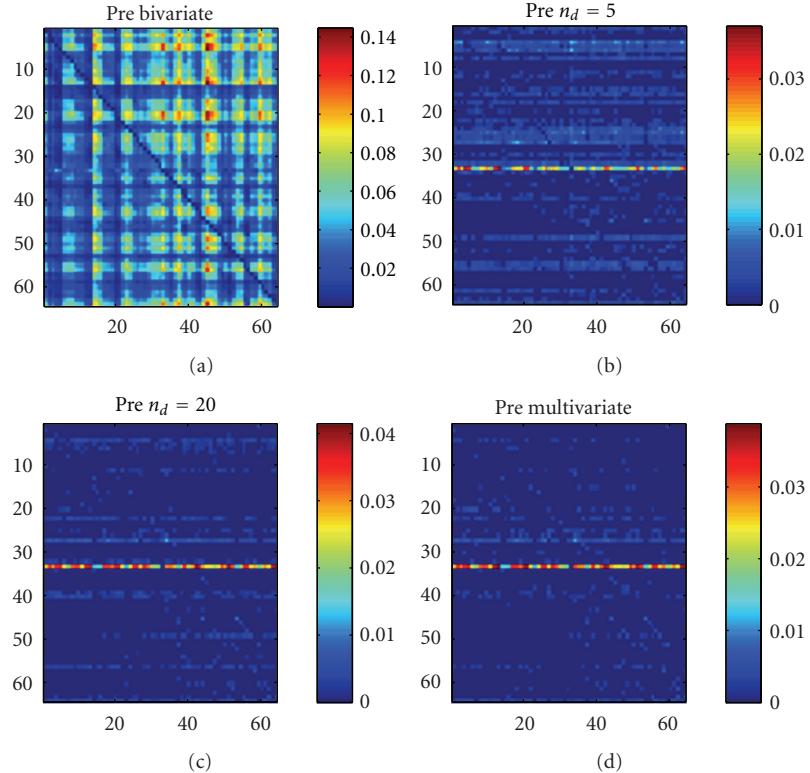


FIGURE 6: The causality analysis of the preictal period. The causality $c(i \rightarrow j)$ corresponds to the row i and the column j . The order is chosen as $m = 6$ according to the AIC criterion. (a) Bivariate analysis. (b) Our approach with $n_d = 5$ conditioning variables. (c) Our approach with $n_d = 20$ conditioning variables. (d) The multivariate analysis.

The two definitions are clearly related by a monotonic transformation:

$$c(\beta \rightarrow \alpha) = -\log[1 - \delta(\beta \rightarrow \alpha)]. \quad (6)$$

Here, we first evaluate the causality $\delta(\beta \rightarrow \alpha)$ using the selection of significative eigenvalues described in [26] to address the problem of overfitting in (5); then we use (6) and express our results in terms of $c(\beta \rightarrow \alpha)$ because it is with this definition that causality is twice the transfer entropy, equal to $I\{x_\alpha; X_\beta | \mathbf{X} \setminus X_\beta\}$, in the Gaussian case [25].

Turning now to the central point of this paper, we address the problem of coping with a large number of variables, when the application of the multivariate Granger causality may

be questionable or even unfeasible, whilst bivariate causality would detect also indirect causalities. Here, we show that conditioning on a small number of variables, chosen as the most informative for the candidate driver variable, is sufficient to remove indirect interactions for sparse connectivity patterns. Conditioning on a large number of variables requires a high number of samples in order to get reliable results. Reducing the number of variables, that one has to condition over, would thus provide better results for small data sets. In the general formulation of the Granger causality, one has no way to choose this reduced set of variables; on the other hand, in the framework of information theory, it is possible to individuate the most informative variables

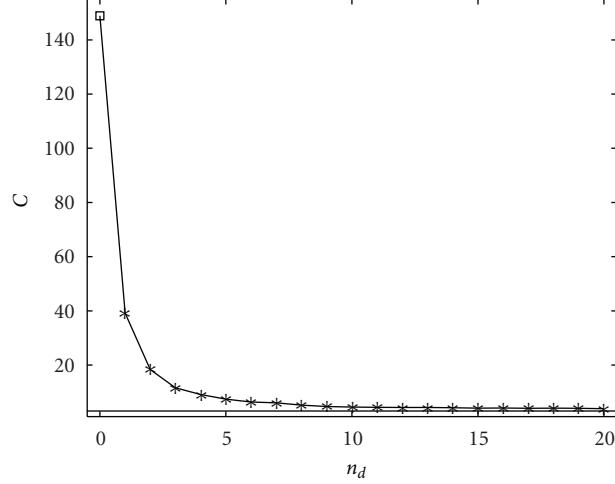


FIGURE 7: Concerning the preictal period, the sum of all causalities is plotted versus the number of conditioning variables.

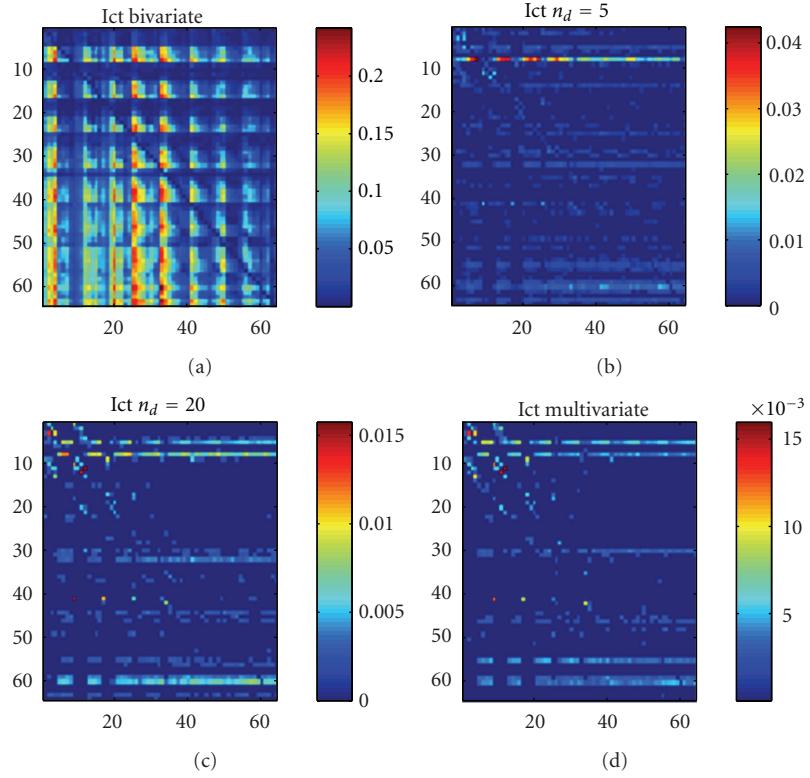


FIGURE 8: The sum of outgoing causality from each electrode in the EEG application, ictal period. (a) Bivariate analysis. (b) Our approach with $n_d = 5$ conditioning variables. (c) Our approach with $n_d = 20$ conditioning variables. (d) The multivariate analysis.

one by one. Once that it has been demonstrated [25] that the Granger causality is equivalent to the information flow between the Gaussian variables, partial conditioning becomes possible for the Granger causality estimation; to our knowledge this is the first time that such approach is proposed.

Concretely, let us consider the causality $\beta \rightarrow \alpha$; we fix the number of variables, to be used for conditioning, equal to n_d . We denote by $\mathbf{Z} = (X_{i_1}, \dots, X_{i_{n_d}})$ the set of the n_d variables, in

$\mathbf{X} \setminus X_\beta$, most informative for X_β . In other words, \mathbf{Z} maximizes the mutual information $I\{X_\beta; \mathbf{Z}\}$ among all the subsets \mathbf{Z} of n_d variables. Then, we evaluate the causality

$$c(\beta \rightarrow \alpha) = \log \frac{\epsilon(x_\alpha | \mathbf{Z})}{\epsilon(x_\alpha | \mathbf{Z} \cup X_\beta)}. \quad (7)$$

Under the Gaussian assumption, the mutual information $I\{X_\beta; \mathbf{Z}\}$ can be easily evaluated, see [25]. Moreover, instead

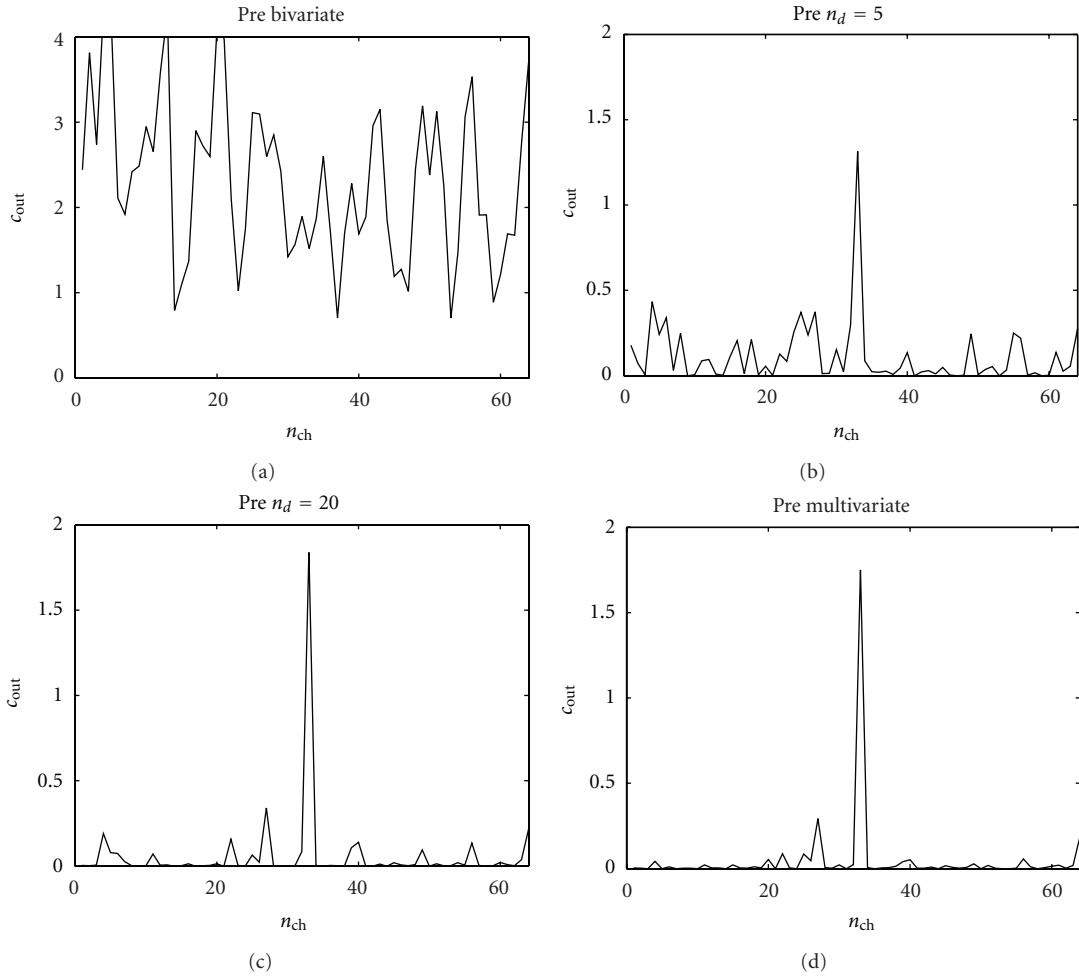


FIGURE 9: The sum of outgoing causality from each electrode in the EEG application, preictal period. (a) Bivariate analysis. (b) Our approach with $n_d = 5$ conditioning variables. (c) Our approach with $n_d = 20$ conditioning variables. (d) The multivariate analysis.

of searching among all the subsets of n_d variables, we adopt the following approximate strategy. Firstly, the mutual information of the driver variable, and each of the other variables, is estimated, in order to choose the first variable of the subset. The second variable of the subsets is selected among the remaining ones as those that, jointly with the previously chosen variable, maximize the mutual information with the driver variable. Then, one keeps adding the rest of the variables by iterating this procedure. Calling Z_{k-1} the selected set of $k-1$ variables, the set Z_k is obtained adding, to Z_{k-1} , the variable, among the remaining ones, with the greatest information gain. This is repeated until n_d variables are selected. This greedy algorithm, for the selection of relevant variables, is expected to give good results under the assumption of sparseness of the connectivity.

3. Results and Discussion

3.1. Simulated Data. Let us consider linear dynamical systems on a lattice of n nodes, with equations, for $i = 1, \dots, n$

$$x_{i,t} = \sum_{j=1}^n a_{ij} x_{j,t-1} + s\tau_{i,t}, \quad (8)$$

where a 's are the couplings, s is the strength of the noise, and τ 's are unit variance i.i.d. Gaussian noise terms. The level of noise determines the minimal amount of samples needed to assess that the structures recovered by the proposed approach are genuine and are not due to randomness as it happens for the standard Granger causality (see discussions in [24, 26]); in particular noise should not be too high to obscure deterministic effects. Firstly we consider a directed tree of 16 nodes depicted in Figure 1; we set a_{ij} equal to 0.9 for each directed link of the graph thus obtained and zero otherwise. We set $s = 0.1$. In Figure 2 we show the application of the proposed methodology to data sets generated by (8), 100 samples long, in terms of quality of the retrieved network, expressed in terms of sensitivity (the percentage of existing links that are detected) and specificity (the percentage of missing links that are correctly recognized as nonexistent). The bivariate analysis provides 100% sensitivity and 92% specificity. However, conditioning on a few variables is sufficient to put in evidence just the direct causalities while still obtaining high values of sensitivity. The full multivariate analysis (obtained as n_d tends to 16) gives here a rather low sensitivity due to the low number of samples. This is a clear example where conditioning on a small number of

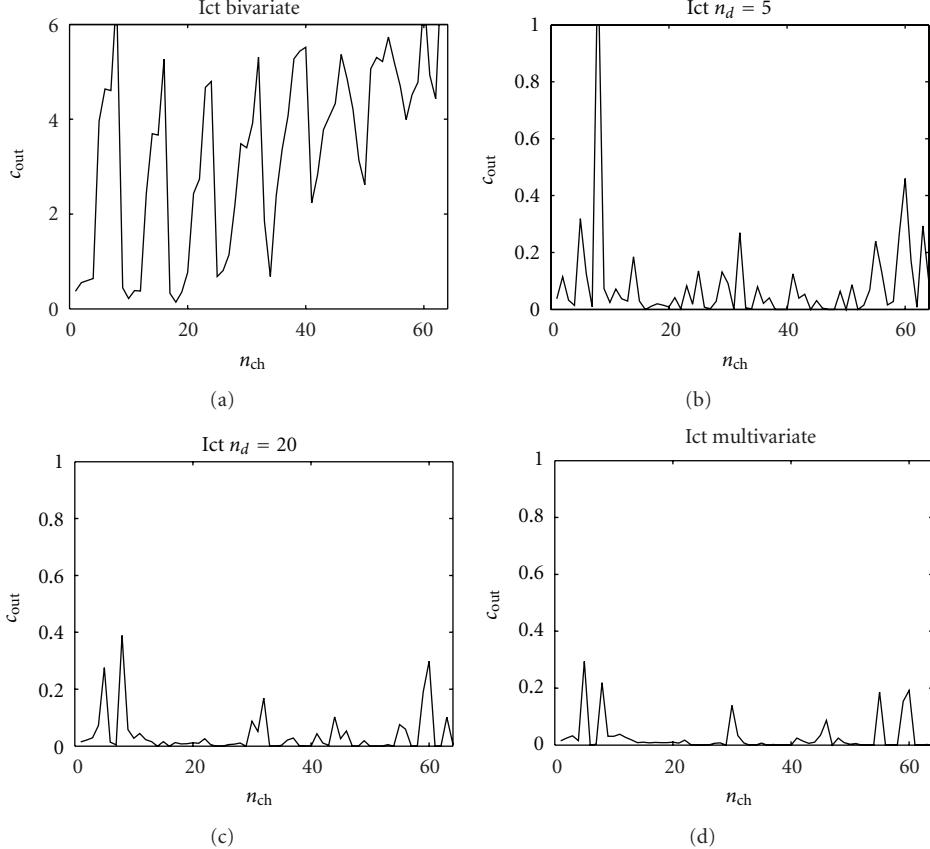


FIGURE 10: The causality analysis of the ictal period. The causality $c(i \rightarrow j)$ corresponds to the row i and the column j . The order is chosen as $m = 6$ according to the AIC criterion. (a) Bivariate analysis. (b) Our approach with $n_d = 5$ conditioning variables. (c) Our approach with $n_d = 20$ conditioning variables. (d) The multivariate analysis.

variables is better than conditioning on all the variables at hand.

As another example, we now fix $n = 34$ and construct couplings in terms of the well-known Zachary data set [27], an undirected network of 34 nodes. We assign a direction to each link, with equal probability, and set a_{ij} equal to 0.015, for each link of the directed graph thus obtained, and zero otherwise. The noise level is set as $s = 0.5$. The network is displayed in Figure 3: the goal is again to estimate this direct-ed network from the measurements of time series on nodes.

In Figure 4 we show the application of the proposed methodology to data sets generated by (8), in terms of sensitivity and specificity, for different numbers of samples. The bivariate analysis detects several false interactions; however, conditioning on a few variables is sufficient to put in evidence just the direct causalities. Due to the sparseness of the underlying graph, we get a result that is very close to the one by the full multivariate analysis; the multivariate analysis here recovers the true network, and indeed the number of samples is sufficiently high. In Figure 5, concerning the stage of selection of variables upon which conditioning is performed, we plot the mutual information gain as a function of the number of variables included n_d : it decreases as n_d increases.

3.2. EEG Epilepsy Data. We consider now a real data set from an 8×8 -electrode grid that was implanted in the cortical surface of the brain of a patient with epilepsy [28]. We consider two 10-second intervals prior to and immediately after the onset of a seizure, called, respectively, the preictal period and the ictal period. In Figure 6 we show the application of our approach to the preictal period; we used the linear causality. The bivariate approach detects many causalities between the electrodes; most of them, however, are indirect. According to the multivariate analysis there is just one electrode that is observed to influence the others, even in the multivariate analysis: this electrode corresponds to a localized source of information and could indicate a putative epileptic focus. In Figure 6 it is shown that conditioning on $n_d = 5$ or $n_d = 20$ variables provides the same pattern corresponding to the multivariate analysis, which thus appears to be robust. These results suggest that the effective connectivity is sparse in the preictal period. As a further confirmation, in Figure 7 we plot the sum of all causalities detected as a function of the number of conditioning variables, for the preictal period; a plateau is reached already for small values of n_d .

In Figure 8 the same analysis is shown w.r.t. the ictal period: in this case conditioning on $n_d = 5$ or $n_d = 20$ variables does not reproduce the pattern obtained with the multivariate approach. The lack of robustness of the causality

pattern w.r.t. n_d seems to suggest that the effective connectivity pattern, during the crisis, is not sparse. In Figures 9 and 10 we show, for each electrode and for the preictal and ictal periods, respectively, the total outgoing causality (obtained as the sum of the causalities on all the other variables). These pictures confirm the discussion above: looking at how the causality changes with n_d may provide information about the sparseness of the effective connectivity.

4. Conclusions

We have addressed the problem of partial conditioning to a limited subset of variables while estimating causal connectivity, as an alternative to full conditioning, which can lead to computational and numerical problems. Analyzing simulated examples and a real data set, we have shown that conditioning on a small number of variables, chosen as the most informative ones for the driver node, leads to results very close to those obtained with a fully multivariate analysis and even better in the presence of a small number of samples, especially when the pattern of causalities is sparse. Moreover, looking at how causality changes with the number of conditioning variables provides information about the sparseness of the connectivity.

References

- [1] A. Barabasi, *Linked*, Perseus, 2002.
- [2] S. Boccaletti, D. U. Hwang, M. Chavez, A. Amann, J. Kurths, and L. M. Pecora, "Synchronization in dynamical networks: evolution along commutative graphs," *Physical Review E*, vol. 74, no. 1, Article ID 016102, 5 pages, 2006.
- [3] Z. Gharhamani, "Learning dynamic bayesian networks," *Lecture Notes in Computer Science*, vol. 1387, pp. 168–197, 1997.
- [4] D. Yu, M. Righero, and L. Kocarev, "Estimating topology of networks," *Physical Review Letters*, vol. 97, no. 18, Article ID 188701, 4 pages, 2006.
- [5] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, no. 1, pp. 1–46, 2007.
- [6] S. L. Bressler and A. K. Seth, "Wiener-Granger causality: a well established methodology," *NeuroImage*, vol. 58, no. 2, pp. 323–329, 2011.
- [7] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance," *Biological Cybernetics*, vol. 85, no. 2, pp. 145–157, 2001.
- [8] K. J. Blinowska, R. Kuś, and M. Kamiński, "Granger causality and information flow in multivariate processes," *Physical Review E*, vol. 70, no. 5, Article ID 050902, 4 pages, 2004.
- [9] A. Seth, "Causal connectivity of evolved neural networks during behavior," *Network*, vol. 16, no. 1, pp. 35–54, 2005.
- [10] A. Roebroeck, E. Formisano, and R. Goebel, "Mapping directed influence over the brain using Granger causality and fMRI," *NeuroImage*, vol. 25, no. 1, pp. 230–242, 2005.
- [11] R. Ganapathy, G. Rangarajan, and A. K. Sood, "Granger causality and cross recurrence plots in rheochoaos," *Physical Review E*, vol. 75, no. 1, Article ID 016211, 6 pages, 2007.
- [12] L. Faes, G. Nollo, and K. H. Chon, "Assessment of granger causality by nonlinear model identification: application to short-term cardiovascular variability," *Annals of Biomedical Engineering*, vol. 36, no. 3, pp. 381–395, 2008.
- [13] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.
- [14] N. Wiener, *The Theory of Prediction*, vol. 1, McGraw-Hill, New York, NY, USA, 1996.
- [15] J. F. Geweke, "Measures of conditional linear dependence and feedback between time series," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 907–915, 1984.
- [16] A. B. Barrett, L. Barnett, and A. K. Seth, "Multivariate Granger causality and generalized variance," *Physical Review E*, vol. 81, no. 4, Article ID 041907, 14 pages, 2010.
- [17] Y. Chen, S. L. Bressler, and M. Ding, "Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data," *Journal of Neuroscience Methods*, vol. 150, no. 2, pp. 228–237, 2006.
- [18] G. Deshpande, S. LaConte, G. A. James, S. Peltier, and X. Hu, "Multivariate granger causality analysis of fMRI data," *Human Brain Mapping*, vol. 30, no. 4, pp. 1361–1373, 2009.
- [19] Z. Zhou, Y. Chen, M. Ding, P. Wright, Z. Lu, and Y. Liu, "Analyzing brain networks with PCA and conditional granger causality," *Human Brain Mapping*, vol. 30, no. 7, pp. 2197–2206, 2009.
- [20] L. Angelini, M. de Tommaso, D. Marinazzo, L. Nitti, M. Pellicoro, and S. Stramaglia, "Redundant variables and granger causality," *Physical Review E*, vol. 3, Article ID 037201, 4 pages, 81.
- [21] D. Marinazzo, W. Liao, M. Pellicoro, and S. Stramaglia, "Grouping time series by pairwise measures of redundancy," *Physics Letters, Section A*, vol. 374, no. 39, pp. 4040–4044, 2010.
- [22] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, pp. 461–464, 2000.
- [23] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 1985.
- [24] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Physical Review Letters*, vol. 100, no. 14, Article ID 144103, 4 pages, 2008.
- [25] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy Are equivalent for gaussian variables," *Physical Review Letters*, vol. 103, no. 23, Article ID 238701, 4 pages, 2009.
- [26] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel-Granger causality and the analysis of dynamical networks," *Physical Review E*, vol. 77, no. 5, Article ID 056215, 9 pages, 2008.
- [27] W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 2, pp. 452–473, 1977.
- [28] M. A. Kramer, E. D. Kolaczyk, and H. E. Kirsch, "Emergent network topology at seizure onset in humans," *Epilepsy Research*, vol. 79, no. 2–3, pp. 173–186, 2008.

Research Article

Quantification of Effective Connectivity in the Brain Using a Measure of Directed Information

Ying Liu and Selin Aviyente

Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA

Correspondence should be addressed to Ying Liu, liuying5@egr.msu.edu

Received 18 November 2011; Revised 20 February 2012; Accepted 14 March 2012

Academic Editor: Luca Faes

Copyright © 2012 Y. Liu and S. Aviyente. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Effective connectivity refers to the influence one neural system exerts on another and corresponds to the parameter of a model that tries to explain the observed dependencies. In this sense, effective connectivity corresponds to the intuitive notion of coupling or directed causal influence. Traditional measures to quantify the effective connectivity include model-based methods, such as dynamic causal modeling (DCM), Granger causality (GC), and information-theoretic methods. Directed information (DI) has been a recently proposed information-theoretic measure that captures the causality between two time series. Compared to traditional causality detection methods based on linear models, directed information is a model-free measure and can detect both linear and nonlinear causality relationships. However, the effectiveness of using DI for capturing the causality in different models and neurophysiological data has not been thoroughly illustrated to date. In addition, the advantage of DI compared to model-based measures, especially those used to implement Granger causality, has not been fully investigated. In this paper, we address these issues by evaluating the performance of directed information on both simulated data sets and electroencephalogram (EEG) data to illustrate its effectiveness for quantifying the effective connectivity in the brain.

1. Introduction

Neuroimaging technologies such as the electroencephalogram (EEG) make it possible to record brain activity with high temporal resolution and accuracy. However, current neuroimaging modalities display only local neural activity rather than large-scale interactions between different parts of the brain. Assessment of the large-scale interdependence between these recordings can provide a better understanding of the functioning of neural systems [1, 2]. Three kinds of brain connectivity are defined to describe such interactions between recordings: anatomical connectivity, functional connectivity, and effective connectivity [2]. Anatomical connectivity is the set of physical or structural connections linking neuronal units at a given time and can be obtained from measurements of the diffusion tensor [3, 4]. Functional connectivity captures the statistical dependence between scattered and often spatially remote neuronal units by measuring their correlations in either time or frequency domain. Effective connectivity describes how one neural

system affects another [2, 4, 5], which can provide information about both the magnitude and the direction of the interaction.

The main approaches used to quantify the effective connectivity between two time series are model-based measures and information-theoretic measures [6]. Granger-causality-based methods and dynamic causal modeling [7] are two widely used model-based measures. Granger causality is a widely used measure to describe the causality between two time series. It defines a stochastic process X causing another process Y if the prediction of Y at the current time point, Y_n , is improved when taking into account the past samples of X . This approach is appealing but gives rise to many questions on how to apply this definition to real data [8]. Granger causality has been mostly applied within a linear prediction framework using a multivariate autoregressive (MVAR) model yielding methods such as directed transfer function (DTF), partial directed coherence (PDC), and directed partial correlation [9–12]. For example, Hesse et al. applied time-varying Granger causality to EEG data and

found that conflict situation generates directional interactions from posterior to anterior cortical sites [10]. Kamiński et al. applied DTF to EEG recordings of human brain during stage 2 sleep and located the main source of causal influence [11]. Schelter et al. employed PDC to EEG recordings from a patient suffering from essential tremor [13]. The extensions of Granger-causality-based methods, such as kernel Granger causality, generalized PDC (gPDC), and extended PDC (ePDC), have also found numerous applications in neuroscience [14–16]. However, Granger causality-based methods, especially those developed from MVAR models, are limited to capturing linear relations or require *a priori* knowledge about the underlying signal models [17]. These approaches may be misleading when applied to signals that are known to have nonlinear dependencies, such as EEG data [18]. DCM, on the other hand, can quantify nonlinear interactions by assuming a bilinear state space model. However, DCM requires *a priori* knowledge about the input to the system [7, 19] and is limited to a network with small size [4]. Thus, a model-free measure detecting both linear and nonlinear relationships is desired.

Information theoretic tools [20–22], such as transfer entropy [20], address the issue of model dependency and have found numerous applications in neuroscience [17, 23, 24]. “Transfer entropy” (TE) proposed by Schreiber computes causality as the deviation of the observed data from the generalized Markov condition and is defined as [20]

$$\begin{aligned} \text{TE}(\mathbf{X} \rightarrow \mathbf{Y}) &= \sum_{y_{n+1}, y_{n-l+1:n}, x_{n-m+1:n}} p(y_{n+1} | y_{n-l+1:n}, x_{n-m+1:n}) \\ &\quad \times \log \frac{p(y_{n+1} | y_{n-l+1:n}, x_{n-m+1:n})}{p(y_{n+1} | y_{n-l+1:n})}, \end{aligned} \quad (1)$$

where m and l are the orders (memory) of the Markov processes \mathbf{X} and \mathbf{Y} , respectively. $p(y_{n+1} | y_{n-l+1:n}, x_{n-m+1:n})$ is the joint probability of random variables $(Y_{n+1}, Y_{n-l+1:n}, X_{n-m+1:n})$, where $Y_{n-l+1:n} = (Y_{n-l+1}, \dots, Y_n)$ and $X_{n-m+1:n} = (X_{n-m+1}, \dots, X_n)$. Sabesan et al. employed TE to identify the direction of information flow for the intracranial EEG data and suggested that transfer entropy plays an important role in epilepsy research [25]. Wibral et al. applied TE to magnetoencephalographic data to quantify the information flow in cortical and cerebellar networks [26]. Vicente et al. extended the definition of TE and measured the information flow from \mathbf{X} to \mathbf{Y} with a general time delay of u , that is, replaced y_{n+1} in the above equation with y_{n+u} , and showed that TE has a better performance in detecting the effective connectivity for nonlinear interactions and signals affected by volume conduction such as real EEG/MEG recordings compared to linear methods [19]. The performance of transfer entropy depends on the estimation of transition probabilities, which requires the selection of order or memory of the Markov processes \mathbf{X} and \mathbf{Y} [25]. “Directed transinformation” (T) introduced by Saito and Harashima [21] measures the information flow from the current sample of one signal to the future samples of another signal given the past samples of both signals. Hinrichs et al. used this measure to analyze

causal interactions in event-related EEG-MEG experiments [17]. However, this measure does not discriminate between totally dependent and independent processes [27]. Recently, directed information proposed by Marko [28] and later reformulated by Massey, Kramer, Tatikonda, and others have attracted attention for quantifying directional dependencies [22, 28–31]. Directed information theory has been mostly aimed towards the study of communication channels with feedback. In recent years, new theoretical developments motivated the use of this measure in quantifying causality between two time series. In particular, Amblard and Michel [31] recently showed how directed information and Granger causality are equivalent for linear Gaussian processes and proved key relationships between existing causality measures and the directed information. Therefore, there has been a growing interest in applying this measure to applications in signal processing, neuroscience, and bioinformatics. For example, it has been successfully used to infer genomic networks [32] and to quantify effective connectivity between neural spike data in neuroscience [31, 33, 34]. In order to detect both linear and nonlinear relationships, in this paper, we propose directed information as a powerful measure to quantify the effective connectivity in the brain.

The theoretical advantages of DI over existing measures have been noted in literature [31, 33, 34]. However, until now the benefits of using DI for capturing the effective connectivity in the brain through neurophysiological data have not been illustrated thoroughly and formally. In addition, because of the relationship between Granger causality and directed information, in this paper, we mainly focus on the comparison between these two measures and investigate the advantage of DI over Granger-causality-based model measures. Theoretical developments only proved the equivalence between these two measures for the case that the time series are distributed as Gaussian in a linear model. However, to date, there has not been much work that compares the actual performance of DI and Granger-causality-based measures for realistic signal models, including both linear and nonlinear interactions. Moreover, most applications of DI to real data have been limited to using either parametric density models for the data or making assumptions about the time dependencies such as assuming a first-order Markov chain and have not considered the difficulties associated with estimating DI from a finite sample size [35]. For complex systems, the computational complexity and the bias of the DI estimator increase with the length of the signal. The main contribution of this paper is to address these issues by evaluating the performance of DI and Granger-causality-based methods under a common framework without making any assumptions about the data distribution. In this paper, we first give a brief introduction to directed information and its computation based on nonparametric estimation methods. We propose a modified time-lagged directed information measure that simplifies the DI computation by reducing the order of the joint entropy terms while still quantifying the causal dependencies. We then evaluate the performance of DI for quantifying the effective connectivity for linear and nonlinear autoregressive models, linear mixing models, single source models, and dynamic chaotic oscillators in

comparison to existing causality measures, in particular with Granger causality. Finally, we apply our method to EEG data to detect the effective connectivity in the brain.

2. Materials and Methods

2.1. Definitions and Notations. In this section, we will first review some common notations and information-theoretic definitions that will be used throughout this paper. Let $\mathbf{X} = X^n = X_{1:n} = (X_1, \dots, X_n)$ be a random process with length n and $p(x_1, \dots, x_n) = p(x^n) = p(x_{1:n})$ be the joint probability of random variables (X_1, \dots, X_n) . $DX^n = X^{n-1} = (0, X_1, \dots, X_{n-1})$ will be used to define the time-delayed version of sequence X^n , which is also equivalent to $X_{1:n-1}$.

Given two continuous random variables X and Y , the mutual information (MI) is defined as follows (All integrals in the paper are from $-\infty$ to $+\infty$ unless otherwise specified.):

$$I(X; Y) = \int \int p(x, y) \log \frac{p(x, y)}{p_x(x)p_y(y)} dx dy, \quad (2)$$

where $p(x, y)$ is the joint probability density function (pdf) of X and Y , and $p_x(x)$, $p_y(y)$ are the marginal pdfs of X and Y , respectively. $I(X; Y) \geq 0$ with equality if and only if X and Y are independent [36]. In information theory, mutual information can be interpreted as the amount of uncertainty about X that can be reduced by observation of Y , or the amount of information Y can provide about X , that is, $I(X; Y) = H(X) - H(X | Y)$. Since $I(X; Y) \geq 0$, $H(X | Y) \leq H(X)$ with equality if and only if X and Y are independent; that is, conditioning reduces entropy [36].

For any three random variables X , Y , and Z , if the conditional distribution of Z depends only on Y and is conditionally independent of X , that is, $p(z | y) = p(z | yx)$, then X , Y , and Z are said to form a Markov chain, denoted by $X \rightarrow Y \rightarrow Z$. In this case, the conditional mutual information between X and Y given Z defined as $I(X; Z | Y) = H(Z | Y) - H(Z | X, Y)$ is equal to 0 [36].

2.2. Directed Information. Mutual information can be extended to random vectors or sequences X^N and Y^N as $I(X^N; Y^N)$, where $I(X^N; Y^N) = H(X^N) - H(X^N | Y^N) = H(Y^N) - H(Y^N | X^N)$. However, mutual information is a symmetric measure and does not reveal any directionality or causality between two random sequences. Massey addressed this issue by defining the directed information from a length N sequence $X^N = (X_1, \dots, X_N)$ to $Y^N = (Y_1, \dots, Y_N)$ [22] as follows:

$$\begin{aligned} DI(X^N \rightarrow Y^N) &= H(Y^N) - H(Y^N | X^N) \\ &= \sum_{n=1}^N I(X^n; Y_n | Y^{n-1}), \end{aligned} \quad (3)$$

where $H(Y^N | X^N)$ is the entropy of the sequence Y^N causally conditioned on the sequence X^N , and $H(Y^N | X^N)$ is defined as

$$H(Y^N | X^N) = \sum_{n=1}^N H(Y_n | Y^{n-1}X^n), \quad (4)$$

which differs from $H(Y^N | X^N) = \sum_{n=1}^N H(Y_n | Y^{n-1}X^n)$ in that X^n replaces X^N in each term on the right-hand side of (4), that is, only the causal influence of the time series \mathbf{X} up to the current time sample n on the process \mathbf{Y} is considered.

An alternative definition of the directed information is proposed by Tatikonda in terms of Kullback-Leibler (KL) divergence [30]. It shows that the difference between mutual information and directed information is the introduction of feedback in the definition of directed information [22, 30, 31]. Mutual information and directed information expressed by KL divergence are written as

$$\begin{aligned} I(X^N; Y^N) &= D_{\text{KL}}(p(x^N, y^N) || p(x^N)p(y^N)), \\ DI(X^N \rightarrow Y^N) &= D_{\text{KL}}(p(x^N, y^N) || \tilde{p}(x^N | y^N)p(y^N)), \end{aligned} \quad (5)$$

where $\tilde{p}(x^N | y^N) = \prod_{n=1}^N p(x_n | x^{n-1}y^{n-1})$ is the feedback factor influenced by the feedback in the system, that is, the probability that the input \mathbf{X} at current time is influenced by the past values of both itself and \mathbf{Y} . If there is no feedback, then $p(x_n | x^{n-1}y^{n-1}) = p(x_n | x^{n-1})$ and $\tilde{p}(x^N | y^N) = p(x^N)$. In fact, $p(x^N, y^N) = \tilde{p}(x^N | y^N)\tilde{p}(y^N | x^N)$, where $\tilde{p}(y^N | x^N) = \prod_{n=1}^N p(y_n | x^n y^{n-1})$ and is defined as the feedforward factor affected by the memory of the system. If the system is memoryless, then $p(y_n | x^n y^{n-1}) = p(y_n | x_n)$.

2.3. Directed Information versus Granger Causality. Granger quantifies causality so that the time series X^N causes Y^N if the variance of the prediction error for \mathbf{Y} at the present time is reduced by including past measurements from \mathbf{X} . Based on Granger's definition of causality, Geweke introduced the Geweke's indices to quantify the causal linear dependencies under Gaussian assumptions [37]. Amblard and Michel proved that the directed information rate and Geweke's indices are equal for Gaussian processes [31] as indicated by

$$DI_\infty(DX^N \rightarrow Y^N) = \frac{1}{2} \log \frac{\varepsilon_\infty^2(Y_N | Y^{N-1})}{\varepsilon_\infty^2(Y_N | Y^{N-1}X^{N-1})} = F_{X^N \rightarrow Y^N}, \quad (6)$$

where DX^N stands for the time-delayed sequence $(0, X_1, \dots, X_{N-1})$ with N being the length of the signal, $DI_\infty(X^N \rightarrow Y^N)$ is the directed information rate; that is, $DI_\infty(X^N \rightarrow Y^N) = \lim_{N \rightarrow \infty} I(X^N; Y_N | Y^{N-1})$, $\varepsilon_\infty^2(Y_N | Y^{N-1}) = \lim_{N \rightarrow \infty} \varepsilon^2(Y_N | Y^{N-1})$ is the asymptotic variance of the prediction residue when predicting Y_N from the observation of Y^{N-1} , and $F_{X^N \rightarrow Y^N}$ refers to the linear feedback measure from random processes X^N to Y^N defined by Geweke [37]. This equality shows that asymptotically the DI rate is equivalent to the gain in information by predicting \mathbf{Y} using the past values of both \mathbf{Y} and \mathbf{X} compared to only using the past samples of \mathbf{Y} , which is similar to the definition of Granger causality. Moreover, Amblard and Michel proved the equality of directed information and Granger's approach for multivariate time series in the case of Gaussian distributions [31].

2.4. Computation of Directed Information. The definition of DI for two length N sequences $X^N = (X_1, \dots, X_N)$ and $Y^N = (Y_1, \dots, Y_N)$ can also be rewritten in terms of the total change of joint entropy or mutual information along time as follows:

$$\begin{aligned} \text{DI}(X^N \rightarrow Y^N) &= \sum_{n=1}^N I(X^n; Y_n | Y^{n-1}) \\ &= \sum_{n=1}^N [H(X^n Y^{n-1}) - H(X^n Y^n)] + H(Y^N) \\ &= \sum_{n=1}^N [I(X^n; Y^n) - I(X^n; Y^{n-1})]. \end{aligned} \quad (7)$$

From the above equations, we can observe that the computation of DI requires the estimation of joint probabilities of high-dimensional random variables over time. If X_n and Y_n are normally distributed, the joint entropy can be estimated based on the covariance matrices. However, for EEG data, the distribution is usually not Gaussian. The nonparametric entropy and mutual information estimators, such as plug-in estimator, m-spacing estimator, and Kozachenko and Leonenko (KL) estimator, have been extensively addressed in literature [38, 39]. In this paper, directed information estimation based on mutual information is used to estimate DI directly from EEG data by using adaptive partitioning method discussed in [39]. However, when the length of the signal increases, the computational complexity, the bias, and the variance of these estimators increase immensely with limited sample sizes. Methods that can reduce the dimension and simplify the computation of DI are needed.

In order to simplify the estimation of DI, we first clarify the connection between the definition of DI used in information theory and the definition as it applies to physical time series. In a physical recording system, if \mathbf{X} starts to influence \mathbf{Y} after p_1 time points or with a delay of p_1 samples, we need to record at least $N + p_1$ time points to obtain N points of the time sequence \mathbf{Y} that have been affected by \mathbf{X} . The directed information rate from time series X^{N+p_1} to Y^{N+p_1} can be defined as [29]. We have

$$\begin{aligned} \text{DI}_\infty(X^{N+p_1} \rightarrow Y^{N+p_1}) &= \lim_{N+p_1 \rightarrow \infty} \frac{1}{N + p_1} \sum_{n=1}^{N+p_1} I(X^n; Y_n | Y^{n-1}) \end{aligned} \quad (8)$$

$$= \lim_{N+p_1 \rightarrow \infty} I(X^{N+p_1}; Y_{N+p_1} | Y^{N+p_1-1}) \quad (9)$$

$$\begin{aligned} &= \lim_{N+p_1 \rightarrow \infty} [H(Y_{N+p_1} | Y^{N+p_1-1}) \\ &\quad - H(Y_{N+p_1} | X^{N+p_1} Y^{N+p_1-1})] \end{aligned} \quad (10)$$

$$\begin{aligned} &= \lim_{N+p_1 \rightarrow \infty} [H(Y_{N+p_1} | Y_{p_1+1:N+p_1-1}) \\ &\quad - H(Y_{N+p_1} | X^{N+p_1} Y_{p_1+1:N+p_1-1})] \end{aligned} \quad (11)$$

$$\begin{aligned} &= \lim_{N+p_1 \rightarrow \infty} [H(Y_{N+p_1} | Y_{p_1+1:N+p_1-1}) \\ &\quad - H(Y_{N+p_1} | X_{1:N} Y_{p_1+1:N+p_1-1})] \end{aligned} \quad (12)$$

$$= \lim_{N+p_1 \rightarrow \infty} I(X_{1:N}; Y_{N+p_1} | Y_{p_1+1:N+p_1-1}) \quad (13)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N I(X^n; Y_{n+p_1} | Y_{p_1+1:n+p_1-1}) \quad (14)$$

$$= \text{DI}_\infty(X_{1:N} \rightarrow Y_{p_1+1:p_1+N}), \quad (15)$$

where (11) comes from the fact that $Y_{1:p_1}$ is independent of Y_{N+p_1} , and (12) is derived using the fact that $X_{N+1:N+p_1}$ has no effect on Y_{N+p_1} because of the time delay p_1 between these two time series. For two physical recordings \mathbf{X} and \mathbf{Y} with length $N + p_1$ and a lag of p_1 , the last equation shows that DI rate for these two time series is equivalent to DI rate for two random processes with length N that are not synchronized in time. In fact, $Y_{p_1+1:p_1+N}$ may be indexed as $Y_{1:N}$ when using the information theoretic indexing, which indexes the signal not according to the physical time point but based on when the receiver receives its first piece of information. Therefore, directed information rate computed by using physical time indices is equivalent to the directed information rate using information theoretic indices for two systems that interact through a time delay. Moreover, when the length of the signal is long enough, the directed information value using both indices will be equivalent.

Once the definition of directed information is extended from random vectors to two physical time series, we propose a modified time-lagged DI to simplify the computation of DI, which is an extension of time-lagged DI proposed for every two samples of X^N and Y^N in [40] to general signal models. As we mentioned before, as the length N of the signal increases, the computational complexity, the bias, and the variance of estimating DI increase immensely with limited sample sizes. In addition, the directed information defined for the physical system is actually a DI with a lag of p_1 samples over a time window with length N . Therefore, an intuitive way to simplify the computation is to apply DI with lag p_1 over a small window. We first give the definition of time-lagged DI for two time series X^N and Y^N with length N at the n th time sample for a block of two time samples with a time delay of p_1 ($n > p_1$):

$$\begin{aligned} \text{DI}_n(X_{n-p_1} X_{n-p_1+1} \rightarrow Y_n Y_{n+1}) &= I(X_{n-p_1}; Y_n) + I(X_{n-p_1} X_{n-p_1+1}; Y_{n+1} | Y_n) \\ &= H(X_{n-p_1}) + H(X_{n-p_1} X_{n-p_1+1} Y_n) + H(Y_n Y_{n+1}) \\ &\quad - H(X_{n-p_1} Y_n) - H(X_{n-p_1} X_{n-p_1+1} Y_n Y_{n+1}), \end{aligned} \quad (16)$$

where $n = p_1 + 1, \dots, N - 1$, p_1 is the time lag between the two time series, and N is the length of the whole time series.

Therefore, the total directed information over the whole time series in terms of the time-lagged DI can be simplified as [40] (the details of the derivation are given in [40]),

$$\begin{aligned} \text{DI}(X^N \rightarrow Y^N) &= \sum_{n=1}^{p_1} I(X^n; Y_n | Y^{n-1}) + \frac{N - p_1}{2(N - p_1 - 1)} \\ &\times \sum_{n=p_1+1}^{N-1} \text{DI}_n(X_{n-p_1} X_{n-p_1+1} \rightarrow Y_n Y_{n+1}). \end{aligned} \quad (17)$$

The time-lagged DI is equivalent to the original definition of DI when p_1 is equal to the actual time delay of the system, the signals \mathbf{X} and \mathbf{Y} follow a single-order model, and Y_n only depends on one past sample of itself, Y_{n-1} . However, these assumptions are not always true. Therefore, we propose the modified time-lagged DI to address these issues.

Consider a general Markov model, where X^N and Y^N are time series with a lag of p_1 and $p(Y_n | X_{1:n-p_1}, Y_{p_1+1:n-1}) = p(Y_n | X_{n-p_2:n-p_1}, Y_{n-p_3:n-1})$, where $p_2 \geq p_1, p_3 \geq 1, p_2$ is the order of the process \mathbf{X} , and p_3 is the order of the process \mathbf{Y} . In this model, it is assumed that \mathbf{X} starts to influence \mathbf{Y} with a delay of p_1 samples, and the order of the model is $p_2 - p_1 + 1$. When the length of the signal N is large enough, then (15) can be further simplified as

$$\begin{aligned} \text{DI}(X^N \rightarrow Y^N) &= \text{DI}(X_{1:N-p_1} \rightarrow Y_{p_1+1:N}) \\ &= \sum_{n=p_1+1}^N I(X_{1:n-p_1}, Y_n | Y_{p_1+1:n-1}) \\ &= \sum_{n=p_1+1}^N [H(Y_n | Y_{p_1+1:n-1}) \\ &\quad - H(Y_n | X_{1:n-p_1} Y_{p_1+1:n-1})]. \end{aligned} \quad (18)$$

Since $p(Y_n | X_{1:n-p_1}, Y_{p_1+1:n-1}) = p(Y_n | X_{n-p_2:n-p_1}, Y_{n-p_3:n-1})$, $X_{1:n-p_2-1} Y_{p_1+1:n-p_3-1} \rightarrow X_{n-p_2:n-p_1}, Y_{n-p_3:n-1} \rightarrow Y_n$ follows a Markov chain. According to Markov chain property,

$$\begin{aligned} &I(X_{1:n-p_2-1} Y_{1:n-p_3-1}; Y_n | X_{n-p_2:n-p_1} Y_{n-p_3:n-1}) \\ &= H(Y_n | X_{n-p_2:n-p_1} Y_{n-p_3:n-1}) \\ &\quad - H(Y_n | X_{1:n-p_1} Y_{p_1+1:n-1}) \\ &= 0, \end{aligned} \quad (19)$$

which means $H(Y_n | X_{n-p_2:n-p_1} Y_{n-p_3:n-1}) = H(Y_n | X_{1:n-p_1} Y_{p_1+1:n-1})$. Therefore,

$$\begin{aligned} \text{DI}(X^N \rightarrow Y^N) &= \sum_{n=p_1+1}^N [H(Y_n | Y_{p_1+1:n-1}) \\ &\quad - H(Y_n | X_{1:n-p_1} Y_{p_1+1:n-1})] \\ &= \sum_{n=p_1+1}^N [H(Y_n | Y_{p_1+1:n-1}) \\ &\quad - H(Y_n | X_{n-p_2:n-p_1} Y_{n-p_3:n-1})] \\ &\leq \sum_{n=p_1+1}^N [H(Y_n | Y_{n-p_3:n-1}) \\ &\quad - H(Y_n | X_{n-p_2:n-p_1} Y_{n-p_3:n-1})] \\ &= \sum_{n=p_1+1}^N I(X_{n-p_2:n-p_1}; Y_n | Y_{n-p_3:n-1}), \end{aligned} \quad (20)$$

where the second equality is using the Markov property, and the inequality comes from the fact that conditioning reduces entropy. For a general Markov model, where X^N and Y^N are stationary statistical processes without instantaneous interaction, such as $p(Y_n | X_{1:n-p_1}, Y_{p_1+1:n-1}) = p(Y_n | X_{n-p_2:n-p_1}, Y_{n-p_3:n-1})$, the modified time-lagged directed information (MDI) is defined as the upper bound of DI:

$$\begin{aligned} \text{MDI}(X^N \rightarrow Y^N) &= \sum_{n=p_1+1}^N I(X_{n-p} \cdots X_{n-1}; Y_n | Y_{n-p} \cdots Y_{n-1}), \end{aligned} \quad (21)$$

where we let $p_1 = 1, p = \max(p_2, p_3)$ to reduce the number of parameters. Note that letting $p_1 = 1$ does not lose any of the information flow compared to using the actual time delay, $p_1 > 1$. The only drawback of letting $p_1 = 1$ is that the computational complexity of estimating the joint entropies increases since the length of the window to compute MDI increases and the dimensionality increases. The main reason why we let $p_1 = 1$ is because estimating the actual value for the delay accurately is not practical when the amount of data is limited. In a lot of similar work such as in [19], different values of p_1 are tested to choose the best one which is not computationally efficient either.

According to (20), modified time-lagged directed information is the upper bound of directed information, that is, $\text{MDI} \geq \text{DI}$. Moreover, MDI is a more general extension of time-lagged DI introduced in our previous work and has two major advantages. First, MDI considers the influence of multiple past samples of \mathbf{Y} on the DI value. Second, it takes into account models with multiple orders; that is, \mathbf{Y} is influenced by different time lags of \mathbf{X} . The modified time-lagged directed information extends the length of the window from 2 to p , which is closer to the actual information flow. When \mathbf{X} and \mathbf{Y} are normally distributed, the computational complexity of the MDI is $O(p^3 N)$ and the computational complexity of the original definition of

DI is $O(N^4)$ (using LU decomposition [41]). Therefore, the computation of MDI is more efficient than that of the original definition of DI.

2.5. Order Selection. For the implementation of MDI, we need to determine the maximum order of the model p . Criterions such as Akaike's final prediction error (FPE) can be used to determine the order of the signal model p . However, this criterion is based on the assumption that the original signal follows a linear AR model and may lead to false estimation of the order when the underlying signal model is nonlinear. Therefore, model-free order selection methods, such as the embedding theorem [42], are needed. For the simplification of computation or parameter estimation, we are only interested in a limited number of variables that can be used to describe the whole system. Suppose we have a time series (X_1, \dots, X_n) , and the time-delay vectors can be reconstructed as $(X_n, X_{n-\tau}, X_{n-2\tau}, \dots, X_{n-(d-1)\tau})$. Projecting the original system to this lower-dimensional state space depends on the choice of d and τ , and the optimal embedding dimension d is related to the order of the model $p = d$ [19]. A variety of measures such as mutual information can be used to determine τ . For discrete time signals, usually the best choice of τ is 1 [43]. To determine d , Cao criterion based on the false nearest neighbor procedure [19] is used to determine the local dimension. The underlying concept of nearest neighbor is that if d is the embedding dimension of a system, then any two points that stay close in the d -dimensional reconstructed space are still close in the $(d+1)$ -dimensional reconstructed space; otherwise, these two points are false nearest neighbors [19, 43]. The choice of d , that is, the model order p , is important for DI estimation. If d is too small, we will lose some of the information flow from \mathbf{X} to \mathbf{Y} . If it is too large, the computational complexity of MDI will be very high, causing the bias and the variance of the estimators to increase.

2.6. Normalization and Significance Test. Since $\text{DI}(X^N \rightarrow Y^N) + \text{DI}(Y^N \rightarrow X^N) = I(X^N; Y^N) + \text{DI}(X^N \rightarrow Y^N || DX^N)$ and $\text{DI}(X^N \rightarrow Y^N) = \text{DI}(DX^N \rightarrow Y^N) + \text{DI}(X^N \rightarrow Y^N || DX^N)$ [29], then

$$\begin{aligned} & \text{DI}(X^N \rightarrow Y^N) + \text{DI}(Y^N \rightarrow X^N) \\ &= \text{DI}(DX^N \rightarrow Y^N) + \text{DI}(X^N \rightarrow Y^N || DX^N) \\ &+ \text{DI}(DY^N \rightarrow X^N) + \text{DI}(Y^N \rightarrow X^N || DY^N). \end{aligned} \quad (22)$$

Therefore,

$$\begin{aligned} & \text{DI}(DX^N \rightarrow Y^N) + \text{DI}(DY^N \rightarrow X^N) \\ &+ \text{DI}(Y^N \rightarrow X^N || DY^N) = I(X^N; Y^N), \end{aligned} \quad (23)$$

where $\text{DI}(Y^N \rightarrow X^N || DY^N) = \text{DI}(X^N \rightarrow Y^N || DX^N)$ indicating the instantaneous information exchange between processes \mathbf{X} and \mathbf{Y} . For a physical system without instantaneous causality, that is, $I(X^N \rightarrow Y^N || DX^N) = 0$, then

$\text{DI}(X^N \rightarrow Y^N) + \text{DI}(Y^N \rightarrow X^N) = I(X^N; Y^N)$ and $0 \leq \text{DI}(X^N \rightarrow Y^N) \leq I(X^N; Y^N) < \infty$. A normalized version of DI, which maps DI to the $[0, 1]$ range, is used for comparing different interactions,

$$\begin{aligned} \rho_{\text{DI}}(X^N \rightarrow Y^N) &= \frac{\text{DI}(X^N \rightarrow Y^N)}{I(X^N; Y^N)} \\ &= \frac{\text{DI}(X^N \rightarrow Y^N)}{\text{DI}(X^N \rightarrow Y^N) + \text{DI}(Y^N \rightarrow X^N)}, \end{aligned} \quad (24)$$

where for a unidirectional system $\mathbf{X} \rightarrow \mathbf{Y}$ with no instantaneous interaction between \mathbf{X} and \mathbf{Y} , $\rho_{\text{DI}}(X^N \rightarrow Y^N) = 1$ and $\rho_{\text{DI}}(Y^N \rightarrow X^N) = 0$; otherwise, if there is no causal relationship between the two signals, the values of $\rho_{\text{DI}}(X^N \rightarrow Y^N)$ and $\rho_{\text{DI}}(Y^N \rightarrow X^N)$ are very close to each other.

In order to test the null hypothesis of noncausality, the causal structure between \mathbf{X} and \mathbf{Y} is destroyed. For each process with multiple trials, we shuffle the order of the trials of the time series \mathbf{X} 100 times to generate new observations \mathbf{X}_m^* , $m = 1, \dots, 100$. In this way, the causality between \mathbf{X} and \mathbf{Y} for each trial is destroyed, and the estimated joint probability changes [44]. We compute the DI for each pair of data $(\mathbf{X}_m^* \text{ and } \mathbf{Y})$. A threshold is obtained at a $\alpha = 0.05$ significance level such that 95% of the directed information for randomized pairs of data ($\text{DI}(\mathbf{X}_m^* \rightarrow \mathbf{Y})$) is less than this threshold. If the DI value of the original pairs of data is larger than this threshold, then it indicates there is significant information flow from \mathbf{X} to \mathbf{Y} .

2.7. Simulated Data. To test the validity and to evaluate the performance of DI for quantifying the effective connectivity, we generate five different simulations. We use these simulation models to compare DI with classical Granger causality (GC) for quantifying causality of both linear and nonlinear autoregressive models, linear mixing models, single source models, and Lorenz systems. The Matlab toolbox developed by Seth is used to compute the GC value in the time domain. GC is also normalized to the $[0, 1]$ range for comparison purposes [45]. The performance of GC depends on the length of the signal, whereas the performance of DI relies on the number of realizations of time series. Therefore, for each simulation, the length of the generated signal for implementing GC is equal to the number of realizations for DI. The significance of DI values are evaluated by shuffling along the trials, while the significance of GC values are evaluated by shuffling along the time series.

Example 1 (Multiple Order Bivariate Linear Autoregressive Model). In this example, we evaluate the performance of DI on a general bivariate linear model,

$$X(n) = \sum_{i=1}^{p_4} \alpha_i X(n-i) + \sigma_x \eta_x(n-1), \quad (25)$$

$$Y(n) = \sum_{i=1}^{p_3} \beta_i Y(n-i) + \gamma \sum_{i=p_1}^{p_2} X(n-i) + \sigma_y \eta_y(n-1). \quad (26)$$

In this bivariate AR model with a delay p_1 and order $p_2 - p_1 + 1$, γ controls the coupling strength between the signals \mathbf{X} and \mathbf{Y} . The initial values of \mathbf{X} and \mathbf{Y} and the noise η_x and η_y are all generated from a Gaussian distribution with mean 0 and standard deviation 1. All coefficients ($\alpha_i, \beta_i, \sigma_x$, and σ_y) are generated from Gaussian distributions with zero mean and unit variance with unstable systems being discarded. To evaluate the performance of directed information, we generate the bivariate model 4096 times with the same parameters but different initial values. γ is varied from 0.1 to 1 with a step size of 0.1, $p_1 = 1$ and $p_2 = p_3 = p_4 = 5$; that is, \mathbf{Y} is influenced by \mathbf{X} through multiple time lags. Without loss of generality, we repeat the simulation 10 times, and average $\text{DI}(\mathbf{X}^N \rightarrow \mathbf{Y}^N)$ and $\text{DI}(\mathbf{Y}^N \rightarrow \mathbf{X}^N)$ over 10 simulations for different γ values. For each simulation, the threshold is evaluated by trial shuffling, and the average threshold is obtained. For GC, the length of the generated signal is chosen as 4096, which is the same as the number of realizations for DI. The GC values in two directions and the corresponding thresholds at the 5% significance level are obtained.

Example 2 (Multiple-Order Bivariate Nonlinear Autoregressive Model). In this example, we evaluate the performance of DI on a general bivariate nonlinear model

$$X(n) = \sum_{i=1}^{p_4} \alpha_i X(n-i) + \sigma_x \eta_x(n-1), \quad (27)$$

$$\begin{aligned} Y(n) = & \sum_{i=1}^{p_3} \beta_i Y(n-i) \\ & + \gamma \sum_{i=p_1}^{p_2} \frac{1}{1 + \exp(b_1 + b_2 X(n-i))} \\ & + \sigma_y \eta_y(n-1). \end{aligned} \quad (28)$$

For this bivariate nonlinear AR model, the setting for the coupling strength γ and the generation of \mathbf{X} , \mathbf{Y} , η_x , η_y , α_i , β_i , σ_x , σ_y , p_1 , p_2 , p_3 , and p_4 are the same as in Example 1. \mathbf{Y} and \mathbf{X} interact nonlinearly through the sigmoid function. Parameters of this function b_1 and b_2 control the threshold level and slope of the sigmoidal curve, respectively. We set $b_1 = 0$ and $b_2 = 50$. DI value and its threshold are averaged over 10 simulations for different γ . The GC values in two directions and the corresponding thresholds at 5% significance level are obtained.

Example 3 (Linear Mixing Model). In this example, we test the effectiveness of DI in inferring effective connectivity when there is linear mixing between these two signals. Linear instantaneous mixing is known to exist in human noninvasive electrophysiological measurements such as EEG or MEG. Instantaneous mixing from coupled signals onto sensor signals by the measurement process degrades signal asymmetry [19]. Therefore, it is hard to detect the causality between the two signals. For unidirectional coupled signal

pairs $\mathbf{X} \rightarrow \mathbf{Y}$ described in (25) to (28), we create two linear mixtures \mathbf{X}_ϵ and \mathbf{Y}_ϵ as follows:

$$\begin{aligned} \mathbf{X}_\epsilon(n) &= (1 - \epsilon)\mathbf{X}(n) + \epsilon\mathbf{Y}(n), \\ \mathbf{Y}_\epsilon(n) &= \epsilon\mathbf{X}(n) + (1 - \epsilon)\mathbf{Y}(n), \end{aligned} \quad (29)$$

where ϵ controls the amount of linear mixing and is varied from 0.05 to 0.45 with a step size of 0.05, and γ is fixed to 0.8 for both models. When $\epsilon = 0.5$, the two signals are identical. Both DI and GC are used to quantify the information flow between \mathbf{X}_ϵ and \mathbf{Y}_ϵ in the two directions.

Example 4 (Single-Source Model). A single source is usually observed on different signals (channels) with individual channel noises [19], which is common in EEG signals due to the effects of volume conduction. In this case, false positive detection of effective connectivity occurs for methods such as Granger causality [46], which means that GC has low specificity. We generate two signals \mathbf{X}_ϵ and \mathbf{Y}_ϵ as follows to test the specificity of DI when there is no significant information flow from one signal to the other signal. We have

$$\begin{aligned} S(n) &= \sum_{i=1}^{p_4} \alpha_i S(n-i) + \eta_S(n), \\ X_\epsilon(n) &= S(n), \\ Y_\epsilon(n) &= (1 - \epsilon)S(n) + \epsilon\eta_Y(n), \end{aligned} \quad (30)$$

where $S(n)$ is the common source generated by an autoregressive model, order $p_4 = 5$, α_i and $\eta_S(n)$ are generated from a Gaussian distribution with mean 0 and standard deviation 1. $S(n)$ is measured on both sensors \mathbf{X}_ϵ and \mathbf{Y}_ϵ . \mathbf{Y}_ϵ is further corrupted by independent Gaussian noise $\eta_Y(n)$ with 0 mean and unit variance. ϵ controls the signal to noise ratio (SNR) in \mathbf{Y}_ϵ and is varied from 0.1 to 0.9 with a step size of 0.1, corresponding to SNR in the range of $-19 \sim 19$ dB.

Example 5 (Nonlinear Dynamic System). In this example, we illustrate the applicability of DI to coupled Lorenz oscillators with a certain delay. The Lorenz oscillator is a three-dimensional dynamic system that exhibits chaotic behavior. Synchronization of two Lorenz systems has been widely investigated for the analysis of EEG data because the dynamic interactions related to the behavior of the cortex can be exemplified by these coupled systems [47]. In the following, we examined two asymmetric coupled Lorenz oscillators (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) as follows [48]:

$$\begin{aligned} \dot{X}_1(t) &= -A(X_1(t) - Y_1(t)), \\ \dot{Y}_1(t) &= RX_1(t) - Y_1(t) - X_1(t)Z_1(t), \\ \dot{Z}_1(t) &= X_1(t)Y_1(t) - BZ_1(t), \\ \dot{X}_2(t) &= -A(X_2(t) - Y_2(t)) + \beta X_1(t - t_p), \\ \dot{Y}_2(t) &= RX_2(t) - Y_2(t) - X_2(t)Z_2(t), \\ \dot{Z}_2(t) &= X_2(t)Y_2(t) - BZ_2(t), \end{aligned} \quad (31)$$

where each equation is a first-order differential equation. $A = 10$, $R = 28$, $B = 8/3$, and $t_p = 0.02$ represents the time delay between two coupled components of these two oscillators, that is, X_1 and X_2 . β corresponds to the coupling strength and is varied from 0.1 to 1 with a step size of 0.2. The differential equations are numerically integrated with a time step of 0.01 using Euler's method [49], corresponding to a delay of 2 time samples between X_1 and X_2 . The initial conditions of these six components are randomly generated from a Gaussian distribution with zero mean and unit variance. We generate 100 samples, and the first 90 samples are discarded to eliminate the initial transients. We compute the information flow in two directions over 10 time points, and the significance of the obtained DI value is verified by trial shuffling.

2.8. Biological Data. In this paper, we examine EEG data from ten undergraduates at Michigan State University drawn from an ongoing study of relationships between the error-related negativity (ERN) and individual differences (Participants for the present analysis were drawn from samples reported on in [50, 51]) such as worry and anxiety. ERN is a brain potential response that occurs following performance errors in a speeded reaction time task [52]. All participants retained for analysis make at least six errors for computation of stable ERNs, as in [53]. Participants complete a letter version of the Eriksen Flanker task [52]. Stimuli are presented on a Pentium R Dual Core computer, using Presentation software (Neurobehavioral systems, Inc.) to control the presentation and timing of stimuli, the determination of response accuracy, and the measurement of reaction times. Continuous electroencephalographic activity is recorded by 64 Ag-AgCl electrodes placed in accordance with the 10/20 system. Electrodes are fitted in a BioSemi (BioSemi, Amsterdam, The Netherlands) stretch-lycra cap. All bioelectric signals are digitized at 512 Hz using ActiView software (BioSemi). For each subject, EEG data are pre-processed by the spherical spline current source density (CSD) waveforms to sharpen event-related potential (ERP) scalp topographies and eliminate volume conduction [54]. In addition, a bandpass filter is used to obtain signals in the theta band. In this study we focus on 33 electrodes corresponding to the frontal, central, and parietal regions of the brain. For each pair of 33 electrodes X and Y for each subject, the effective connectivity is quantified by computing the modified time-lagged DI over 70 trials and a model order of p in the theta band. The model order or the length of the time window p is determined by the Cao Criterion. We also apply Granger causality to the same data and compare its performance with directed information.

Previous work indicates that there is increased synchronization associated with ERN for the theta frequency band (4–8 Hz) and ERN time window 25–75 ms after the response for error responses (ERN) in the anterior cingulate cortex (ACC), in particular between the lateral prefrontal cortex (IPFC) and medial prefrontal cortex (mPFC) [55]. In this paper, we wish to verify these existing findings using the proposed DI measure and to further infer the directional causality underlying these dependencies.

3. Results and Discussion

In this section, we first evaluate the effectiveness of directed information on quantifying both linear and nonlinear causal relationships through simulated data and compare the performance of directed information with GC. We then apply the directed information to real EEG data to reveal the pairwise information flow in the brain.

3.1. Simulated Data

Example 1 (Multiple-Order Bivariate Linear Autoregressive Model). In this example, the DI value in two directions averaged across 10 simulations with different γ is shown in Figure 1(a). The performance of GC is shown in Figure 1(b). The estimated order of the model is $p = 5$, which is in accordance with the simulation model. γ controls the coupling strength between X and Y . We observe that $DI(X^N \rightarrow Y^N)$ is significant for all values of γ . On the contrary, $DI(Y^N \rightarrow X^N)$ is less than the threshold, which indicates the acceptance of the null hypothesis that there is no significant causal information flow from Y to X . Since GC uses a linear autoregressive framework for quantifying causality; in this example, GC detects the causality relationship between X and Y successfully; that is, the information flow from X to Y is significant for all γ while it is insignificant for the opposite direction. It is also interesting to note that GC and DI exhibit similar behavior across different values of γ , indicating the equivalency of the two measures for linear Gaussian signal models.

Example 2 (Multiple-Order Bivariate Nonlinear Autoregressive Model). In this example, the performance of DI and GC for the nonlinear autoregressive model in (27) and (28) averaged across 10 simulations with different γ are evaluated as shown in Figure 2. The estimated order of the model is 5. We observe that when γ is less than 0.3, the coupling strength between X and Y is weak and the DI value in both directions is not significant. As γ increases, $DI(X^N \rightarrow Y^N)$ increases and becomes significant. $DI(Y^N \rightarrow X^N)$ decreases with increasing γ and is still less than the threshold as expected. The results indicate increased unidirectional information flow from X to Y with increasing γ and show that detecting the information flow in nonlinear processes is more difficult especially when the coupling strength is low. GC fails to detect the information flow from X to Y for all γ . Since GC is implemented in a linear framework, the estimated order and the model itself do not match with the nonlinearity of the signal. Therefore, it cannot detect nonlinear causality.

Example 3 (Linear Mixing Model). For this example, the DI value and GC value averaged across 10 simulations with changing linear mixing coefficient ϵ for both linear and nonlinear AR models are shown in Figure 3. The estimated order of the model is 5 as before. When $\epsilon = 0.5$, the two observed mixing signals are identical, and we expect to see no significant information flow in the two directions. We observe that, for the linear AR model, directed information detects the causality between X_ϵ and Y_ϵ when ϵ is smaller

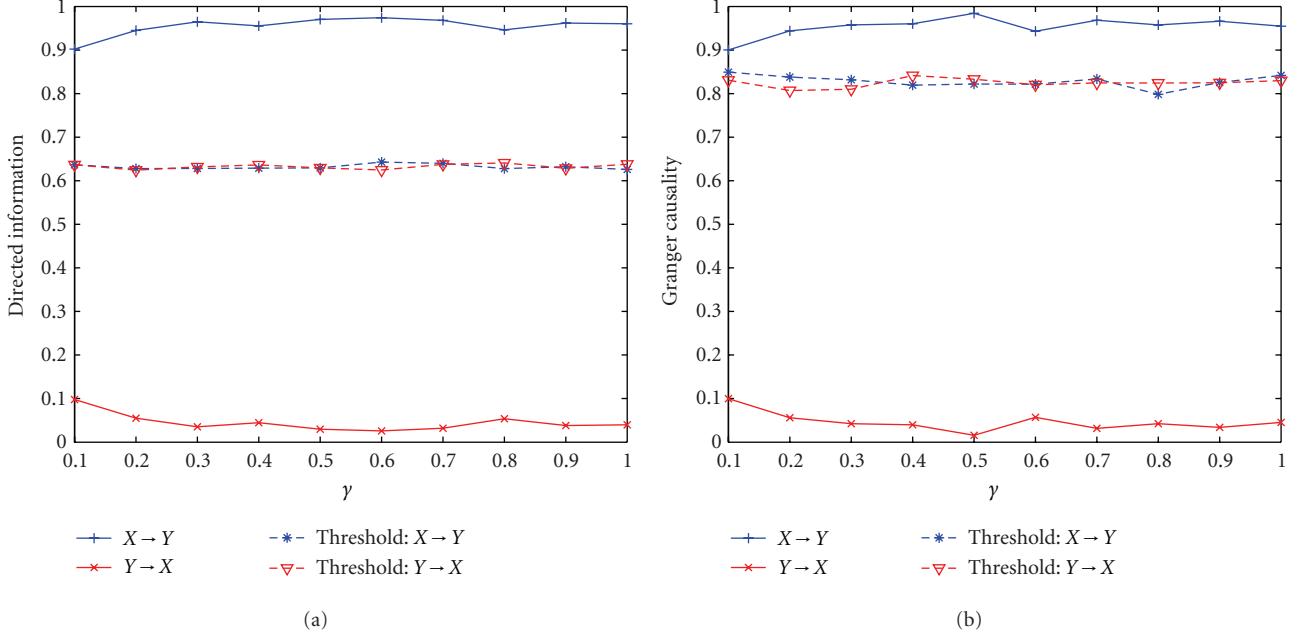


FIGURE 1: Application of directed information and Granger causality to bivariate linear autoregressive model. (a) Directed information with different γ . (b) Granger causality with different γ .

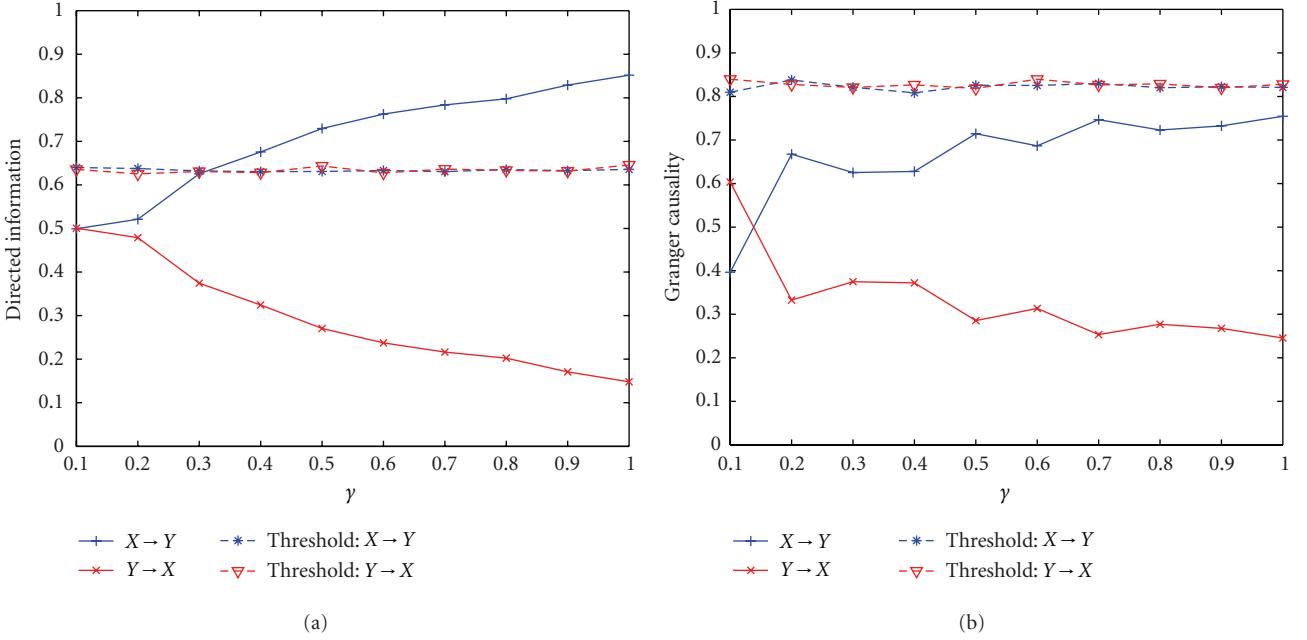


FIGURE 2: Application of directed information and Granger causality to bivariate nonlinear autoregressive model. (a) Directed information with different γ . (b) Granger causality with different γ .

than 0.4. When ϵ is larger than 0.4, the causality between X_ϵ and Y_ϵ is hard to detect because of the strong mixing; that is, X_ϵ and Y_ϵ are almost identical, and the information flow in both directions becomes insignificant. Compared to DI, GC only detects the causality from X_ϵ to Y_ϵ when the mixing is weak ($\epsilon < 0.2$), indicating that GC is more vulnerable

to linear mixing. It is probably due to the fact that GC is sensitive to the mixture of signals, and the assumed signal model does not match with the original signal [46]. For the nonlinear AR model, DI fails to detect causality when ϵ is larger than 0.1, which indicates that linear mixing of nonlinear source models makes it harder to detect effective

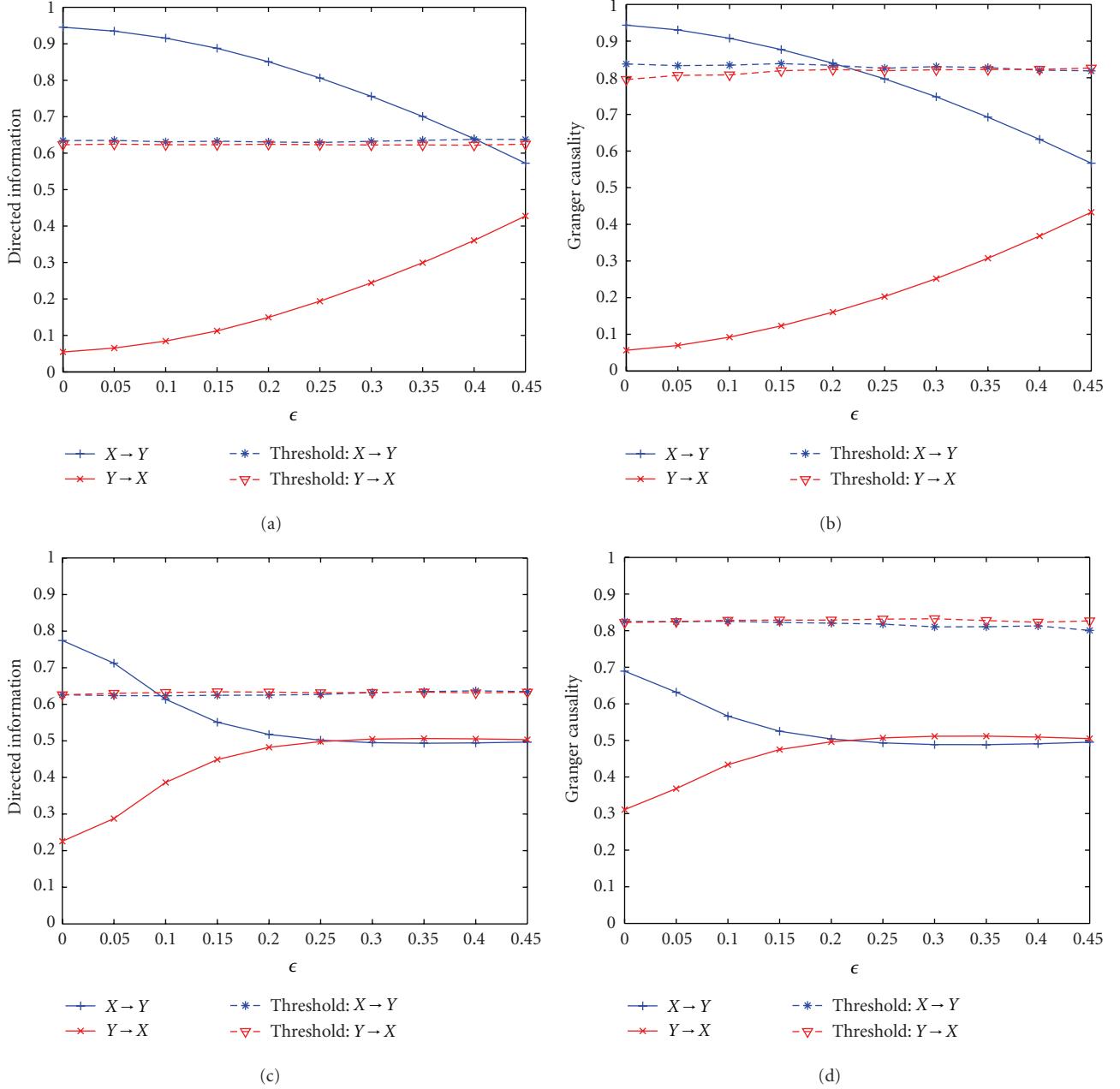


FIGURE 3: Application of directed information and Granger causality to linear mixing for both linear and nonlinear autoregressive models. (a) Directed information with different ϵ for the linear mixing of linear AR model. (b) Granger causality with different ϵ for the linear mixing of linear AR model. (c) Directed information with different ϵ for the linear mixing of nonlinear AR model. (d) Granger causality with different ϵ for the linear mixing of nonlinear AR model.

connectivity compared to mixing of linear source models. On the other hand, GC fails to detect any causality even when $\epsilon = 0$ since it cannot detect nonlinear interactions.

Example 4 (Single-Source Model). We use the single source model to test the specificity of DI. The DI value and GC value averaged across 100 simulations for changing ϵ for a single source model are shown in Figure 4. The estimated order of the model is 5 as before. In addition, the false positive

rate using both DI and Granger causality with increasing ϵ is also calculated. We observe that the information flow in two directions using DI is less than the threshold for all values of ϵ , which indicates the acceptance of the null hypothesis that there is no significant causal information flow from X to Y or Y to X . Note that DI is normalized by the mutual information. For a common source model, the instantaneous information exchange between X and Y contributes mostly to the mutual information between X and Y . Thus, according to (23), $DI(DX^N \rightarrow Y^N)$ and $DI(DY^N \rightarrow X)$ normalized by

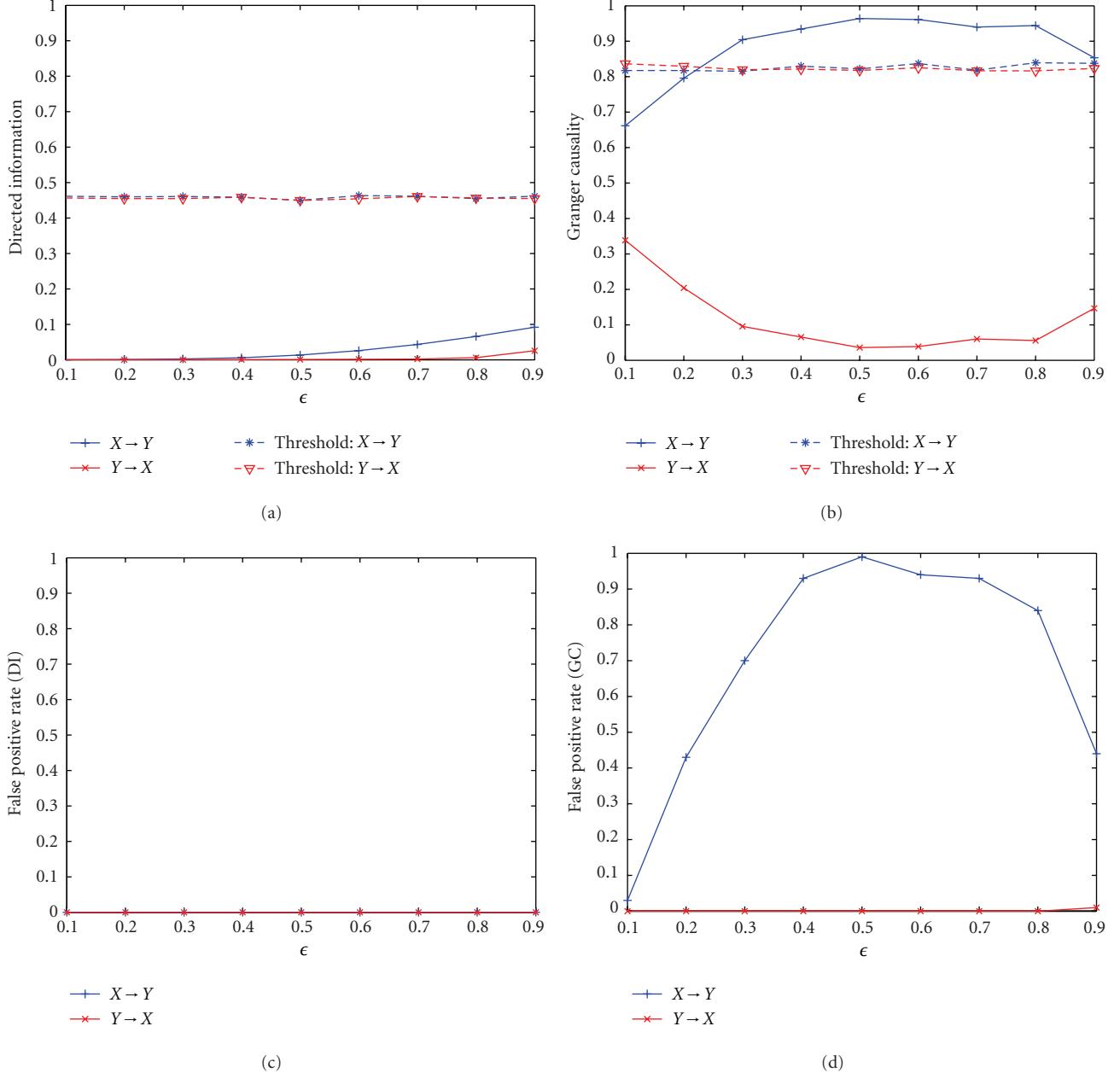


FIGURE 4: Application of directed information and Granger causality to single source model. (a) Directed information with different ϵ for the single source model. (b) Granger causality with different ϵ for the single source model. (c) False positive rate for directed information with different ϵ for the single source model. (d) False positive rate for Granger causality with different ϵ for the single source model.

mutual information are close to 0 and less than the threshold from the randomized data pairs. The false positive rate of DI is 0 for all ϵ . Therefore, DI is able to discriminate between instantaneous mixing from actual causality and is very robust to noise. For GC, when ϵ is small (<0.2) or large (>0.9), the value of GC is less than or very close to the threshold in both directions thus indicating that there is no causal information flow between the two processes. However, GC fails to accept the null hypothesis when ϵ is between 0.3 to 0.9 and detects a nonexisting effective connectivity. GC reaches its maximum value when $\epsilon = 0.5$. This is due to the fact that GC is close to

0 when two processes X and Y are independent or identical, that is, when $\epsilon = 1$ and $\epsilon = 0$. Based on the definition of GC, the prediction of Y at the current time point will not be improved by taking into account the past samples of X for these processes [26]. Therefore, as ϵ increases from 0 to 0.5, X becomes the most different from Y ; therefore, it can provide more new information about Y and the GC increases. As ϵ increases from 0.5 to 1, X becomes independent of Y , and the GC decreases. The false positive rate of GC is not equal to 0 for all values of ϵ , which indicates that it has lower specificity compared to DI. Therefore, GC is not robust to

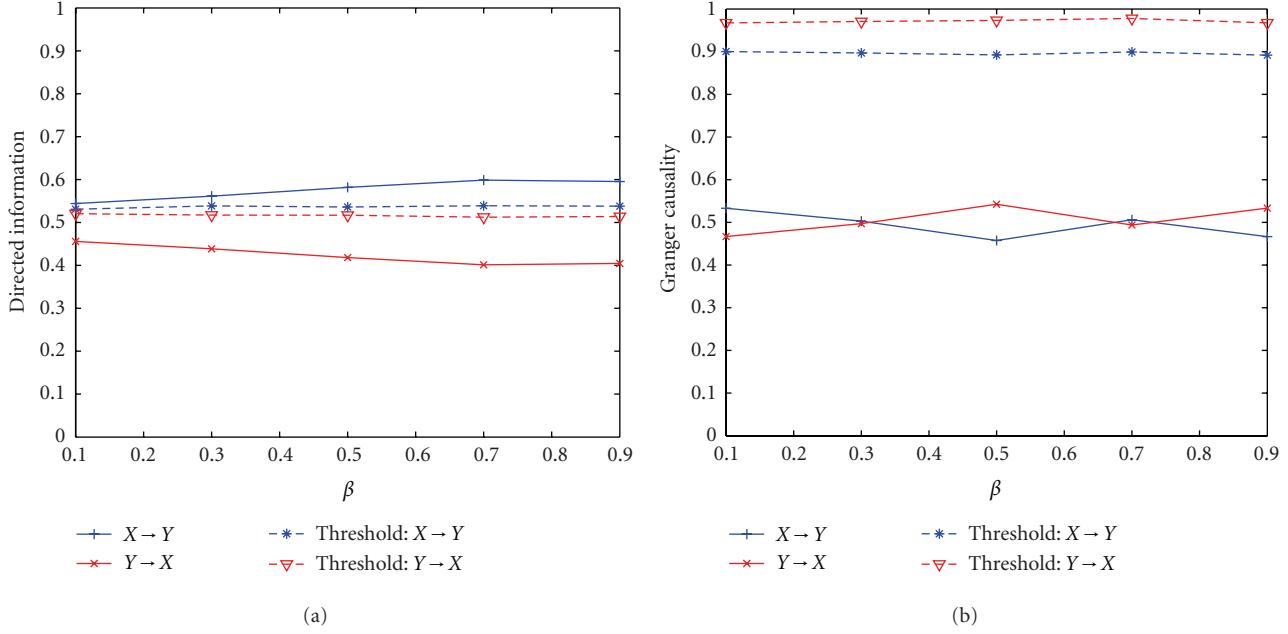


FIGURE 5: Application of directed information and Granger causality to two asymmetric coupled Lorenz oscillators. (a) Directed information with different β . (b) Granger causality with different β .

the effect of a common source and may infer false positive effective connectivity. This simulation indicates that DI is more sensitive and discriminative about the information flow patterns in the presence of volume conduction, which means it is a more promising method to capture the effective connectivity for real EEG data.

Example 5 (Nonlinear Dynamic System). In this example, the DI values and GC values between X_1 and X_2 of two asymmetric coupled Lorenz systems are computed with coupling strength β being set from 0.1 to 1. The estimated order of the model is 3. Though this is larger than the actual model order, our method will not lose any information except for the increased computational complexity. The results are shown in Figure 5. The results show that DI values from X_1 to X_2 increase with the coupling strength β and are significant for all values of β . In addition, there is no significant causal information flow from X_2 to X_1 . Therefore, DI can effectively detect the causality in a nonlinear dynamic system. On the contrary, GC cannot detect any significant information flow for all β values. It is due to the fact that the model selected for implementing GC is not consistent with the dynamic characteristics of the system.

3.2. EEG Data. Previous work indicates that there is increased information flow associated with ERN for the theta frequency band (4–8 Hz) and ERN time window 25–75 ms for error responses compared to correct responses in particular between mPFC and iPFC regions [55]. In addition, Cavanagh et al. have shown that there is increased synchronization for error trials between electrode pairs, such as FCz-F5 and FCz-F6, compared to the synchrony between FCz-CP3 and FCz-CP4 [56]. The DI and GC values for each pair

of electrodes averaged over 10 subjects are computed over a time window of 53 time points (100 ms). The estimated order of the model for each electrode pairs is 3. In order to control the error rates for multiple hypothesis testing for all pairs of electrodes, the method proposed by Genovese et al. is used in this paper [57]. To implement this procedure, for two electrodes with time series \mathbf{X} and \mathbf{Y} , we first shuffle the order of the trials of \mathbf{X} 100 times to generate new observations \mathbf{X}_m^* , $m = 1, \dots, 100$. The P value of $\text{DI}(\mathbf{X} \rightarrow \mathbf{Y})$ is obtained by comparing it with DI values from randomized pairs of data $\text{DI}(\mathbf{X}_m^* \rightarrow \mathbf{Y})$, $m = 1, \dots, 100$. We then obtain the threshold P_r for all P values ($33 \times 33 \times 10$) by controlling the FDR bound q as 0.05. For $\text{DI}(\mathbf{X} \rightarrow \mathbf{Y})$, if the P value is less than P_r , then the directed information flow from \mathbf{X} to \mathbf{Y} is significant; otherwise, it is not significant. Electrode pairs between which the information flow is significant in at least one of the ten subjects are shown in Figure 6(b). We also test the significance of Granger causality in the same way. When the FDR is controlled at 0.05, the information flow between electrode pairs is significant if the P -value of DI or GC is less than 0.01. Electrode pairs that have significant causality relationship using both measures are shown in Figure 6. In Figures 6(a) and 6(c), each small circle shows the directed information and Granger causality from a particular electrode to other electrodes. In Figures 6(b) and 6(d), each small circle shows electrode pairs that have significant causality relationship. The results indicate that DI detects strong information flow from the frontal region (e.g., F5 and F6) to the frontal-central region (e.g., FC2 and FCz) corresponding to the lateral prefrontal cortex (iPFC) and medial prefrontal cortex (mPFC). In addition, the central-parietal region (e.g., CPz, CP1, and CP2) around the midline, corresponding to the motor cortex, has strong

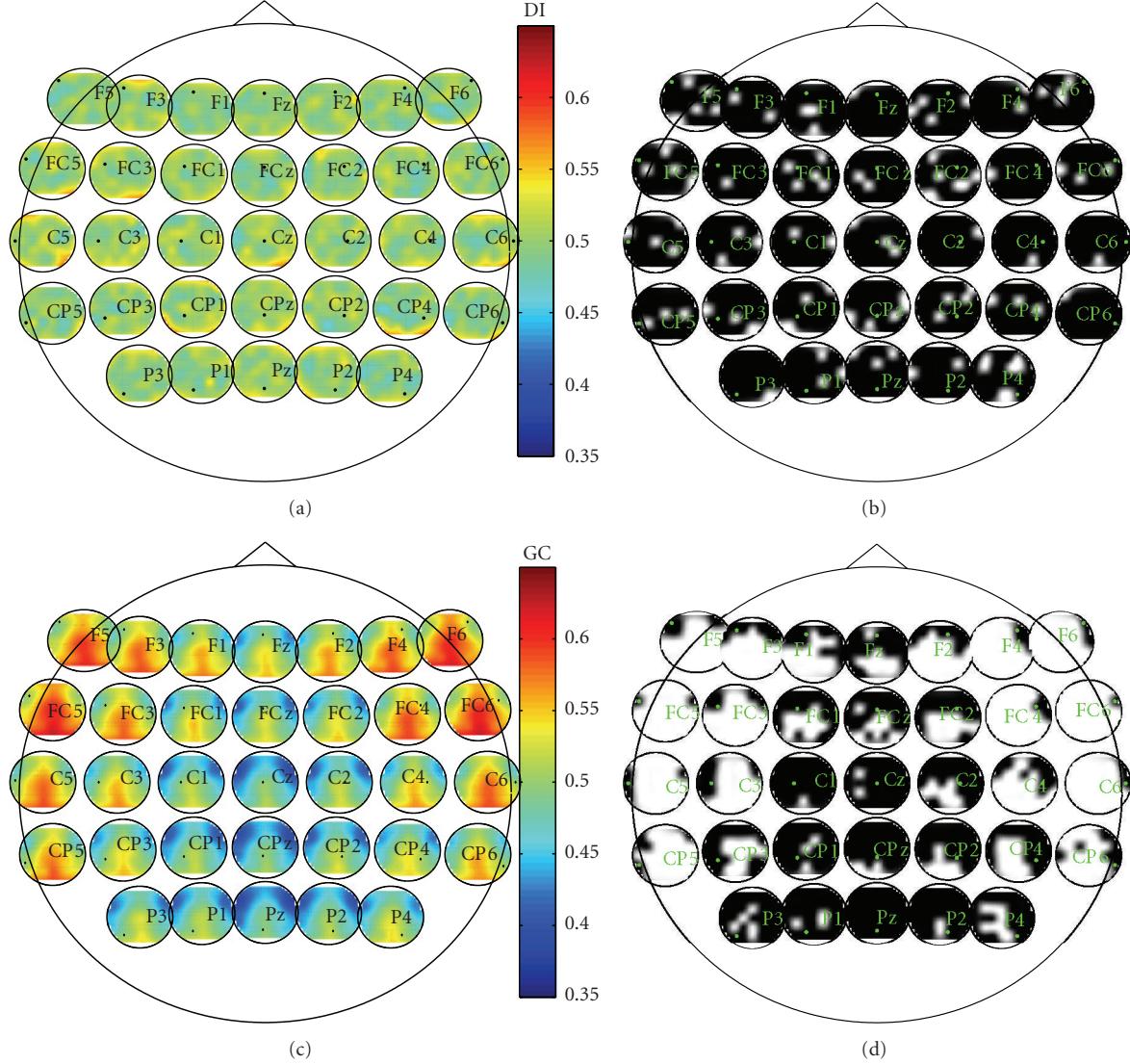


FIGURE 6: Application of directed information and Granger causality to EEG data. (a) Pairwise directed information. (b) Electrode pairs with significant DI values. (c) Pairwise Granger causality. (d) Electrode pairs with significant GC values. For (b) and (d), green dots indicate the location of the particular node, and white regions correspond to significant information flow from that particular electrode to other electrodes.

influence on the central and frontal regions (e.g., FCz and F6) since this is a speeded response task involving the motor cortex. The details of the significant electrode interactions are shown in Table 1. These results are aligned with the previous work in [56], which shows that error processing is controlled by the communication between the lateral prefrontal cortex and medial prefrontal cortex. When GC is applied to the same data, the information flow pattern around the midline is similar to the DI. However, the information flow from the lateral prefrontal cortex to the rest of the brain is significant. On one hand, the similar patterns of connectivity using both measures verify the validity of proposed DI computation algorithm. On the other hand, GC shows significance over a wide region of the brain especially in the lateral areas compared to DI, which may be due

to GC's low specificity to volume conduction in the form of a common source. Previous work and our simulation in Example 4 have indicated that Granger-causality-based measures may infer erroneous effective connectivity in the case of the common source as seen in EEG data [19, 46]. However, without ground truth, we cannot confirm that some links reported as significant by GC are spurious and due to volume conduction in a conclusive manner, but the results from DI agree more with the suggestions in [56], that most of the increase in connectivity during cognitive control, that is, ERN, should be between medial prefrontal cortex and lateral prefrontal cortex, compared to the results of GC. Therefore, DI is more sensitive and discriminative about the information flow patterns compared to GC for real neurophysiological data.

TABLE 1: Electrode pairs in the region of interest with significant DI values.

From	To	From	To	From	To
F5	F1 FC2 CPz CP4 P3	C5	F6 FC5 Cz CP4	P3	P4
F3	FC3 CP4	C3	FC2 C6 P1	P1	F2 C6 CP2
F1	C1 Cz Pz	C1	FC1 C6	Pz	F5 F4 FCz
FZ	F5	CZ	F5 C2 CP4	P2	FC4 C5
F2	FC3 FC6 C5 CP1	C2	FC6	P4	F3 F4 FC3 FC2 FC4 Pz P2
F4	F6 C4	C4	P2		
F6	F2 FC3 FCz Cz	C6	Pz		
FC5	Fz C3 C2 CP6 CP5		Cz C4 CP3		
FC3	CP1	CP3	C5 CPz P4		
FC1	F4 FC3 C2 CP1 CP4	CP1	F6 FCz P3		
FCZ	C3 CP1	CPz	FC6 C6 CP5 CP4 P1		
FC2	F3 C1 C6 CP2 CP4 P3	CP2	F6 FCz FC4 CP1		
FC4	C5	CP4	FC5 FCz C4		
FC6	C5 C4 CP1	CP6	F5		

4. Conclusions

In this paper, we illustrated the advantages of a new directed information measure over Granger-causality-based measures for quantifying the effective connectivity in the brain. In order to illustrate the advantages of this measure, first, we applied directed information measure to identify the causality relationships for both linear and nonlinear AR models, linear mixing models, single source models, and Lorenz systems and compare its performance with Granger causality. Directed information is shown to be more effective in detecting the causality of different systems compared to Granger causality. We then applied the directed information measure on EEG data from a study containing the error-related negativity to infer the information flow patterns between different regions. The results showed that the directed information measure can capture the effective connectivity in the brain between the mPFC and lPFC areas as predicted by previous work.

Directed information, as a model-free measure, is able to detect both linear and nonlinear causality relationships between two signals. However, other model-free entropy-based measures would also detect effective connectivity such as transfer entropy and directed transinformation. Directed transinformation introduced by Saito measures the information flow from the current sample of one signal to the future samples of another signal given the past samples of both signals but does not discriminate between totally dependent and independent processes. Transfer entropy and directed information are very closely related to each other. Transfer entropy quantifies the information gained at

each time step by measuring the deviation of the observed data from the generalized Markov condition. Therefore, the definition of transfer entropy implicitly assumes a stationary Markov process [31]. Compared to transfer entropy, directed information quantifies the sum of information obtained over the whole time series [58] and does not make any assumptions about the underlying signal model. Thus, theoretically, the original definition of directed information can apply to any signal models. In real applications, in order to simplify the computation of directed information, we usually make certain assumptions about the underlying signal model such as the modified time-lagged DI proposed in this paper, which basically assumes a stationary Markov process similar to transfer entropy. In addition, Amblard and Michel proved that, for a stationary process, directed information rate can be decomposed into two parts, one of which is equivalent to the transfer entropy when $l = m = n$ in (1) and the other to the instantaneous information exchange rate [31]. In another words, for a physical system without instantaneous interactions between its subsystems, the rate of these two measures, directed information and transfer entropy, is equivalent asymptotically as the length of the signal goes to infinity.

There are still remaining issues with the implementation of directed information. First, the performance of directed information relies on accurate estimation from limited sample sizes that introduces bias to the estimated values. This problem can be addressed by either using parametric density models or improving existing mutual information and entropy estimators. Recently, Zhao et al. proposed an universal algorithm to estimate directed information for stationary ergodic processes by using sequential probability assignment, which may be used to improve the effective connectivity results discussed in this paper [59]. Second, the performance of directed information relies on the selection of the model order. If the order of the model is too small, it will lose the information from X to Y. If it is too large, the computational complexity is very high. In addition to classical embedding dimension determination methods such as the Cao criterion used in this paper, Faes et al. proposed a sequential procedure to determine the embedding dimension of multivariate series [60]. This method is based on an information-theoretic technique and shows promising performances for various signal models, which may be extended to DI computation in the future. Third, directed information does not discriminate between direct and indirect interactions among multivariate time series. However, this is not a shortcoming of DI since DI does not assume any particular signal interaction model: bivariate or multivariate. Similar to other information theoretic measures, such as mutual information, whether the particular measure can identify interactions between multiple processes depends on how the measure is applied. For example, in the case of mutual information, though the original definition is for two random processes X and Y, it is possible to extend it to multiple processes [61]. Similarly, we can apply DI over multiple processes using conditional directed information such as the definition given by Kramer. We address this issue in a previous paper [34] by using conditional directed

information and develop an algorithm to infer the actual network. Similarly, GC originally is defined for two time series that a stochastic process X causing another process Y if the prediction of Y at the current time point, Y_n , is improved when taking into account the past samples of X . However, in application it has been extended to multiple processes through the use of multivariate AR models. Future work will focus on the comparison of these two measures in a multivariate setting.

Acknowledgments

The authors would like to thank Dr. Jason Moser from the Department of Psychology at Michigan State University for sharing his EEG data. This work was in part supported by the National Science Foundation under Grants No. CCF-0728984 and CAREER CCF-0746971.

References

- [1] S. L. Bressler, "Large-scale cortical networks and cognition," *Brain Research Reviews*, vol. 20, no. 3, pp. 288–304, 1995.
- [2] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag, "Organization, development and function of complex brain networks," *Trends in Cognitive Sciences*, vol. 8, nos. 9, pp. 418–425, 2004.
- [3] M. A. Koch, D. G. Norris, and M. Hund-Georgiadis, "An investigation of functional and anatomical connectivity using magnetic resonance imaging," *NeuroImage*, vol. 16, no. 1, pp. 241–250, 2002.
- [4] K. J. Friston, "Functional and effective connectivity: a review," *Brain Connectivity*, vol. 1, no. 1, pp. 13–36, 2011.
- [5] K. J. Friston, "Functional and effective connectivity in neuroimaging: a synthesis," *Human Brain Mapping*, vol. 2, no. 1-2, pp. 56–78, 1994.
- [6] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progress in Neurobiology*, vol. 77, no. 1-2, pp. 1–37, 2005.
- [7] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, 2003.
- [8] C. W. J. Granger, "Testing for causality. A personal viewpoint," *Journal of Economic Dynamics and Control*, vol. 2, no. 1, pp. 329–352, 1980.
- [9] L. A. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.
- [10] W. Hesse, E. Möller, M. Arnold, and B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *Journal of Neuroscience Methods*, vol. 124, no. 1, pp. 27–44, 2003.
- [11] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biological Cybernetics*, vol. 85, no. 2, pp. 145–157, 2001.
- [12] W. Mader, D. Feess, D. Saur et al., "Investigating multivariate systems using directed partial correlation," *International Journal of Bioelectromagnetism*, vol. 12, no. 1, pp. 21–25, 2001.
- [13] B. Schelter, M. Winterhalder, M. Eichler et al., "Testing for directed influences among neural signals using partial directed coherence," *Journal of Neuroscience Methods*, vol. 152, no. 1-2, pp. 210–219, 2006.
- [14] D. Marinazzo, W. Liao, H. Chen, and S. Stramaglia, "Nonlinear connectivity by Granger causality," *NeuroImage*, vol. 58, no. 2, pp. 330–338, 2010.
- [15] L. Leistritz, T. Weiss, J. Ionov, K. J. Bär, W. H. R. Miltner, and H. Witte, "Connectivity analysis of somatosensory evoked potentials to noxious intracutaneous stimuli in patients with major depression," *Methods of Information in Medicine*, vol. 49, no. 5, pp. 484–491, 2010.
- [16] L. Faes and G. Nollo, "Extended causal modeling to assess partial directed coherence in multiple time series with significant instantaneous interactions," *Biological Cybernetics*, vol. 103, no. 5, pp. 387–400, 2010.
- [17] H. Hinrichs, T. Noesselt, and H. J. Heinze, "Directed information flow—a model free measure to analyze causal interactions in event related EEG-MEG-experiments," *Human Brain Mapping*, vol. 29, no. 2, pp. 193–206, 2008.
- [18] F. Lopes da Silva, J. P. Pijn, and P. Boeijinga, "Interdependence of EEG signals: linear versus nonlinear associations and the significance of time delays and phase shifts," *Brain Topography*, vol. 2, no. 1-2, pp. 9–18, 1989.
- [19] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—a model-free measure of effective connectivity for the neurosciences," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 45–67, 2011.
- [20] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, pp. 461–464, 2000.
- [21] Y. Saito and H. Harashima, *Recent Advances in EEG and EMG Data Processing*, Elsevier, Amsterdam, The Netherlands, 1981.
- [22] J. Massey, "Causality, feedback and directed information," in *Proceedings of the International Symposium on Information Theory and Its Applications*, pp. 27–30, 1990.
- [23] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [24] M. Lungarella and O. Sporns, "Mapping information flow in sensorimotor networks," *PLoS Computational Biology*, vol. 2, no. 10, pp. 1301–1312, 2006.
- [25] S. Sabesan, L. B. Good, K. S. Tsakalis, A. Spanias, D. M. Treiman, and L. D. Iasemidis, "Information flow and application to epileptogenic focus localization from intracranial EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 3, pp. 244–253, 2009.
- [26] 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," *Progress in Biophysics and Molecular Biology*, vol. 105, no. 1-2, pp. 80–97, 2011.
- [27] M. Al-Khassawneh and S. Aviyente, "The relationship between two directed information measures," *IEEE Signal Processing Letters*, vol. 15, pp. 801–804, 2008.
- [28] H. Marko, "The bidirectional communication theory—a generalization of information theory," *IEEE Transactions on Communications*, vol. 21, no. 12, pp. 1345–1351, 1973.
- [29] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 4–21, 2003.
- [30] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 323–349, 2009.
- [31] P. O. Amblard and O. J. J. Michel, "On directed information theory and Granger causality graphs," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 7–16, 2011.
- [32] A. Rao, A. O. Hero III, D. J. States, and J. D. Engel, "Using directed information to build biologically relevant influence,"

- in *Proceedings of the Computational Systems Bioinformatics*, pp. 145–156, 2007.
- [33] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, “Estimating the directed information to infer causal relationships in ensemble neural spike train recordings,” *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 17–44, 2011.
- [34] Y. Liu and S. Aviyente, “Information theoretic approach to quantify causal neural interactions from EEG,” in *Proceedings of the 44th IEEE Asilomar Conference on Signals, Systems and Computers (ASILOMAR ’10)*, pp. 1380–1384, November 2010.
- [35] P. Mathai, N. C. Martins, and B. Shapiro, “On the detection of gene network interconnections using directed mutual information,” in *Proceedings of the Information Theory and Applications Workshop (ITA ’07)*, pp. 274–283, February 2007.
- [36] T. M. Cover, J. A. Thomas, and J. Wiley, *Elements of Information Theory*, Wiley Online Library, 1991.
- [37] J. Geweke, “Measurement of linear dependence and feedback between multiple time series,” *Journal of the American Statistical Association*, vol. 77, no. 378, pp. 304–313, 1982.
- [38] E. G. Miller, “A new class of entropy estimators for multidimensional densities,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 297–300, April 2003.
- [39] G. A. Darbellay and I. Vajda, “Estimation of the information by an adaptive partitioning of the observation space,” *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1315–1321, 1999.
- [40] Y. Liu and S. Aviyente, “Time-lagged directed information,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’11)*, pp. 3864–3867, 2011.
- [41] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, 3rd edition, 2007.
- [42] B. Schelter, M. Winterhalder, and J. Timmer, *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, VCH Verlagsgesellschaft mbH, 2006.
- [43] L. Cao, “Practical method for determining the minimum embedding dimension of a scalar time series,” *Physica D*, vol. 110, no. 1-2, pp. 43–50, 1997.
- [44] G. Pipa and S. Grün, “Non-parametric significance estimation of joint-spike events by shuffling and resampling,” *Neurocomputing*, vol. 52, pp. 31–37, 2003.
- [45] A. K. Seth, “A MATLAB toolbox for Granger causal connectivity analysis,” *Journal of Neuroscience Methods*, vol. 186, no. 2, pp. 262–273, 2010.
- [46] G. Nolte, A. Ziehe, V. V. Nikulin et al., “Robustly estimating the flow direction of information in complex physical systems,” *Physical Review Letters*, vol. 100, no. 23, Article ID 234101, 2008.
- [47] M. Breakspear, “Nonlinear phase desynchronization in human electroencephalographic data,” *Human Brain Mapping*, vol. 15, no. 3, pp. 175–198, 2002.
- [48] W. Michiels and H. Nijmeijer, “Synchronization of delay-coupled nonlinear oscillators: an approach based on the stability analysis of synchronized equilibria,” *Chaos*, vol. 19, no. 3, Article ID 033110, 2009.
- [49] J. C. Butcher and J. Wiley, *Numerical Methods for Ordinary Differential Equations*, vol. 2, Wiley Online Library, 2003.
- [50] J. S. Moser, H. S. Schroder, C. Heeter, T. P. Moran, and Y.-H. Lee, “Mind your errors: evidence for a neural mechanism linking growth mindset to adaptive post-error adjustments,” *Psychological Science*, vol. 22, no. 12, pp. 1484–1489, 2011.
- [51] J. S. Moser, T. Moran, and A. Jendrusina, “Parsing relationships between dimensions of anxiety and action monitoring brain potentials in female undergraduates,” *Psychophysiology*, vol. 49, no. 1, pp. 3–10, 2012.
- [52] B. A. Eriksen and C. W. Eriksen, “Effects of noise letters upon the identification of a target letter in a nonsearch task,” *Perception and Psychophysics*, vol. 16, no. 1, pp. 143–149, 1974.
- [53] D. M. Olvet and G. Hajcak, “The stability of error-related brain activity with increasing trials,” *Psychophysiology*, vol. 46, no. 5, pp. 957–961, 2009.
- [54] J. Kayser and C. E. Tenke, “Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks,” *Clinical Neurophysiology*, vol. 117, no. 2, pp. 348–368, 2006.
- [55] S. Aviyente, E. M. Bernat, W. S. Evans, and S. R. Sponheim, “A phase synchrony measure for quantifying dynamic functional integration in the brain,” *Human Brain Mapping*, vol. 32, no. 1, pp. 80–93, 2011.
- [56] J. F. Cavanagh, M. X. Cohen, and J. J. B. Allen, “Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring,” *Journal of Neuroscience*, vol. 29, no. 1, pp. 98–105, 2009.
- [57] C. R. Genovese, N. A. Lazar, and T. Nichols, “Thresholding of statistical maps in functional neuroimaging using the false discovery rate,” *NeuroImage*, vol. 15, no. 4, pp. 870–878, 2002.
- [58] J. Lizier, *The local information dynamics of distributed computation in complex systems*, Ph.D. dissertation, University of Sydney, 2010.
- [59] L. Zhao, H. Permuter, Y. H. Kim, and T. Weissman, “Universal estimation of directed information,” in *Proceedings of the IEEE International Symposium on Information Theory (ISIT ’10)*, pp. 1433–1437, June 2010.
- [60] L. Faes, G. Nollo, and A. Porta, “Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique,” *Physical Review E*, vol. 83, no. 5, Article ID 051112, 2011.
- [61] W. Zhao, E. Serpedin, and E. R. Dougherty, “Inferring connectivity of genetic regulatory networks using information-theoretic criteria,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5, no. 2, pp. 262–274, 2008.

Research Article

An Analytical Approach to Network Motif Detection in Samples of Networks with Pairwise Different Vertex Labels

**Christoph Schmidt,¹ Thomas Weiss,² Christian Komusiewicz,³
Herbert Witte,¹ and Lutz Leistritz¹**

¹ Institute of Medical Statistics, Computer Sciences and Documentation, Jena University Hospital, Friedrich Schiller University Jena, Bachstrasse 18, 07743 Jena, Germany

² Department of Biological and Clinical Psychology, Friedrich Schiller University Jena, Am Steiger 3, Haus 1, 07743 Jena, Germany

³ Institut für Softwaretechnik und Theoretische Informatik, TU Berlin, Fakultät IV, Elektrotechnik und Informatik, Sekr. FR 6-1, Franklinstrasse 28/29, 10587 Berlin, Germany

Correspondence should be addressed to Christoph Schmidt, christoph.schmidt@mti.uni-jena.de

Received 10 November 2011; Revised 17 January 2012; Accepted 6 February 2012

Academic Editor: Mingzhou Ding

Copyright © 2012 Christoph Schmidt et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Network motifs, overrepresented small local connection patterns, are assumed to act as functional meaningful building blocks of a network and, therefore, received considerable attention for being useful for understanding design principles and functioning of networks. We present an extension of the original approach to network motif detection in single, directed networks without vertex labeling to the case of a sample of directed networks with pairwise different vertex labels. A characteristic feature of this approach to network motif detection is that subnetwork counts are derived from the whole sample and the statistical tests are adjusted accordingly to assign significance to the counts. The associated computations are efficient since no simulations of random networks are involved. The motifs obtained by this approach also comprise the vertex labeling and its associated information and are characteristic of the sample. Finally, we apply this approach to describe the intricate topology of a sample of vertex-labeled networks which originate from a previous EEG study, where the processing of painful intracutaneous electrical stimuli and directed interactions within the neuromatrix of pain in patients with major depression and healthy controls was investigated. We demonstrate that the presented approach yields characteristic patterns of directed interactions while preserving their important topological information and omitting less relevant interactions.

1. Introduction

Many processes and systems have a network structure that consists of interacting units which can be represented as a graph. Accordingly, analysis from a graph theory perspective has recently become a focus of research as unique insights are obtained into the working and organization of various complex systems. For example, in the study of cellular signaling pathways associated with cancer it was revealed that the activity of p53, a central tumor suppressor that regulates many different genes, can only be understood by considering associated tangled signaling networks in their entirety and the position of p53 integration within these networks, instead of considering interactions of p53 with single network components [1]. In synthetic biology, “network thinking” is crucial

for the understanding and assembly of biological modules that are used in engineering cellular machines to perform tasks such as producing drugs or acting as biosensors that detect toxic compounds [2, 3]. In epidemiology, considerable effort has been directed to examining mechanisms by which the topology of networks of contacts between individuals affects the spreading of diseases in order to find ways to predict and control the propagation of infections [4]. The progress made in modern network theory has also led to new applications in the neurosciences that attempt to find explanations for previously inadequately understood higher level brain processes [5]. Topological properties of anatomical and functional connectivity networks have been studied to obtain understanding on the organization of cortical areas.

It has been suggested that brain systems exhibit a small-world topology which implicates simultaneous and well-balanced segregation and integration of information processing and results in minimization of wiring costs for economical brain performance [6, 7]. Disturbances of this evolutionary optimized topology of cortical networks have been reported to alter functional connectivity and thereby cause neuropsychiatric disorders such as Alzheimer's disease [8], schizophrenia [9, 10], or epilepsy [11] that are often described as disconnection syndromes [12–14]. Pathological abnormalities in cortical network organization may be quantified by network measures which might act as useful diagnostic markers. An overview of measures that quantify global and local network topology can be found for example in [15–17].

Network motifs, the subject of this publication, constitute an exceptional influential measure of local network topology that enables a detailed description of overrepresented local patterns of interconnections [18, 19]. Subsequently, these overrepresented subnetworks may be linked with a potential functional contribution to the global functionality of the entire network. The functionality of a system is to some extent enclosed or encoded in the topology of its representing network; as a consequence, it is assumed that individual networks (or at least networks of a certain type) possess characteristic combinations of recurring small, connected subnetworks that act as functional meaningful building blocks or as elementary computational circuits for information processing [18, 20, 21]. The importance of the functional contribution of a subnetwork is assumed to be reflected in an overrepresented, nonrandom and perhaps conserved occurrence of it in its network. According to this, a functional constraint for subnetworks correlating with their nonrandom appearance is robustness to small perturbations in order to enable robust network performance, especially in biological networks [22]. Network motif detection has been directly adopted into a variety of different research fields; interesting results have been obtained by its application to study structure-function relationships and design principles in networks from various domains such as protein-protein interaction networks, the World Wide Web, electronic circuits, synaptic neuronal networks, and transcriptional gene regulation networks [18–20, 23, 24]. A slightly modified variant of motif detection was used to investigate structural motifs and the instances of functional motifs contained within them in the context of anatomic brain networks of macaque visual cortex, macaque cortex, and cat cortex [25].

The original network motif detection approach attempts to find significant frequent subnetworks in one single-directed network with (usually many) unlabeled vertices that are indistinguishable from each other. Original network motif detection basically consists of three computationally expensive subtasks.

- (1) Exhaustively enumerating [18] or sampling (estimating) [26, 27] the number of occurrences of each subnetwork of size k (a subnetwork induced by a vertex set of k vertices) in the input network. This quantity is affected by the kind of vertex and edge overlap one allows for counting different matches

of a subnetwork [28, 29]. Typically, one allows for nonidentical counting (arbitrary overlaps) of subnetworks where the downward closure property does not hold. This dramatically increases the number of subnetwork occurrences in a network compared to counting only edge-disjoint subnetworks, that is, subnetworks that do not share any edges. Therefore, even in comparably small networks, the number of subnetwork occurrences is potentially large due to its exponential increase with the size of the input network. Moreover, the number of k -subnetworks in a network grows very fast with k . In order to avoid impractical running times and difficulties with assessing functional roles of larger subnetworks, the size parameter k for subnetworks is usually chosen to be 3 or 4.

- (2) The second subtask in network motif detection encompasses determining graph isomorphism for grouping found subnetworks into equivalence classes. It is believed that graph isomorphism cannot be solved in polynomial time. Several algorithms for solving graph isomorphism with miscellaneous performance in practice have been presented [30].
- (3) The last subtask is assessing statistical significance of subnetwork occurrences. Subnetworks that occur in significantly large numbers in the input network as compared to their occurrence in a large set of null model random networks are accepted to be motifs. The comparison of a network with a set of associated random networks should reveal deviations of network properties such as the number of subnetwork occurrences from randomness. Therefore, the underlying random graph model has to be chosen carefully, because it is this model that specifies the notion of randomness. Hence, it has to strike a balance between preserving functional constraints and characteristics of the input network while at the same time comprising random edge patterns so that at best no subnetwork appearance is being favored [31]. The commonly employed random graph model preserves the incoming and outgoing degree sequence—an important characteristic of single vertices—of the input network and the associated random networks are usually generated either by the configuration model (stubs matching) algorithm [32–35] or by an Markov chain Monte Carlo edge rewiring (switching) algorithm [18, 32, 33, 36, 37]. A subnetwork occurrence is defined to be significant if it occurs a certain multiple of standard deviations more often in the input network than would be expected in the set of random networks. This is expressed by the z -score which relates the count of a subnetwork in the input network to the mean and the standard deviation of its count in the set of random networks [25, 26, 28]. Making use of z -scores for assigning statistical significance to subnetwork occurrences is flawed by the unsafe assumption being made that subnetwork occurrences follow a normal distribution [28] and it

was shown that this is not always the case [38]. Statistical significance of subnetwork occurrences might also be assessed by computing whether the probability that a subnetwork occurs an equal or greater number of times in a random network than in the input network is lower than a cutoff value [18, 19].

We present a novel approach to network motif detection that differs from the original approach and avoids some of its limitations. In this approach, we intend to detect subnetworks that are motifs for a sample of directed networks where each network possesses the same pairwise different vertex labels. Furthermore, we intend to not discard these vertex labels but rather preserve the functional important topological information associated with them. Topological information has already been used in previous studies to visualize spatio-temporal connectivity structures [39–41]. Another advantage of this approach is to analytically compute the statistical significance of subnetwork counts, which would save considerable computation time since no random network ensembles have to be generated and no subnetwork counts have to be obtained from them. Motifs that are obtained by this analysis would yield a description of locatable and characteristic interaction patterns of a sample of networks and moreover could be used as a distinguishing characteristic to reveal sample-specific differences in network topology. We demonstrate that this approach may be applied to investigate networks that model pain processing in a group of patients with major depression (MD) and a group of healthy controls (HCs) in order to acquire deeper insight into the intertwined relationship between pain and depression where many details are poorly understood. It is known that chronic pain and major depression are correlated since depression is a common comorbidity of chronic pain and often chronic pain is an additional symptom of depressed patients [42, 43]. It has been confirmed by some studies that thresholds for acute painful stimulation are lower in depressed patients than in healthy controls [44, 45], whereas other studies found the opposite, namely, increased thresholds in depressed patients [42, 46–49]. The physiological basis for pain perception, pain processing, and the sensitivity to painful stimuli of depressed patients remains unclear. It is hypothesized that in depressed patients the processing of painful stimuli in the so-called neuromatrix of pain [50] and consequently the effective connectivity might be altered [47].

We denote the networks from the samples we investigate in the present study by effective connectivity networks (ECNs). The network data originates from a previous study in which we used effective connectivity analysis to investigate the processing of moderately painful intracutaneous electrical stimuli and directed interactions within the neuromatrix of pain in both groups, MD and HCs, by means of frequency selective generalized partial directed coherence (gPDC) [51]. The intricacy of the connectivity patterns in ECNs does not allow for immediate interpretation. In order to overcome the inability to qualitatively describe the intricate wiring patterns found in ECNs and to shed light on elementary directed interactions in both groups we use our novel approach as a filter that detects labeled network motifs and omits less

important interactions. We demonstrate that in this way we gain interesting new insights into the relationship between chronic pain and depression, which is currently inadequately understood.

2. Materials and Methods

The present study directly follows up on the EEG experiments and the connectivity analysis published in [51] and is based on the same materials. For the sake of completeness, a short sketch of the baseline characteristics is given in this section.

2.1. Subjects. Eighteen patients (10 women, 8 men) with major depression (mean age \pm standard deviation: 38.9 ± 15.5 years) and 18 sex- and age-matched healthy control subjects (39.3 ± 14.8 years) participated in this study. Patients were treated in a specialized psychiatric ward for mood disorders. Major depression was established by a staff psychiatrist according to DSM IV criteria using a structured interview, and the Beck depression inventory (BDI) was also administered. BDI scores of patients ranged from 19 to 48 (29.4 ± 9.7); scores of control subjects were all below five (2.1 ± 1.5). All subjects were right handed. Nine patients were treated with antidepressive medication (5 patients received selective serotonin reuptake inhibitors SSRI; 4 patients norepinephrine and serotonin reuptake inhibitors NaSRI) while the remaining participants did not receive any medication. One patient and two controls had to be excluded from the experiments because they could not follow the test protocol.

2.2. Connectivity Analysis. All subjects were electrically stimulated intracutaneously at the tip of the middle fingers of both the right and the left hand. The intensity level was adjusted between $10 \mu\text{A}$ and $1 \mu\text{A}$. Stimuli consisted of a bipolar rectangular pulse of 10 ms duration. Participants were requested to rate each electrical stimulus on a scale ranging from 0 to 6 (0 = no sensation; 1 = just perceived, not painful; 2 = clearly perceived, but not painful; 3 = low pain; 4 = moderate pain; 5 = strong pain, but tolerable; 6 = unbearable pain) [52, 53]. The pain threshold was defined as the intensity yielding a sensation described as a sharp painful pinprick, corresponding to a rating of "3." The EEG was recorded continuously during the electrical stimulation from 60 electrodes, referenced to Cz, using a standard EEG cap (Easy Cap, Falk Minow Services, Germany) based on an extended International 10–20 system. Finally, nine electrodes F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4 (re-referenced to linked ears) that are situated above some of the important regions of pain processing, attention, and depression (frontal, central, and parietal brain regions) were used. Eye movement and muscle activity artifact contaminated single trial somatosensory evoked potentials (SEPs) were excluded, which resulted in an exclusion of three data sets since there were not enough artifact-free trials left for a reliable connectivity analysis. In order to compare the pre- and post-stimulus condition, signal sections of 700 ms duration were extracted pre-(700 ms before onset to the onset of stimulus, i.e., -700 ms

to 0 ms) as well as post-stimulus onset (from stimulus onset to 700 ms after stimulus onset, i.e., 0 ms to 700 ms). These signal sections provided the data basis for the connectivity analysis. To assess the effective connectivity between each directed pair of the nine electrodes, the generalized partial directed coherence (gPDC) [54] was applied. The frequency range of interest for the SEP analysis was determined to be in the delta-, theta- and the alpha-bands (1 to 13 Hz) since the signal power is mainly situated in this frequency range. For a consolidated analysis, the gPDCs of the corresponding frequencies were pooled to one quantity by averaging with respect to the frequency range of interest. Thus, for each of the 72 possible directed interactions, one gPDC value results each. Finally, the effective connectivity we are interested in is given by significantly increased gPDC values. A detailed description of the entire procedure may be found in [51].

2.3. Effective Connectivity Networks. In this study, we refine the view of this effective connectivity data by examining effective connectivity from a different perspective: we model each participant's directed interactions, which are given by significant gPDC values, as effective connectivity networks (ECNs). The topology of ECNs consequently represents a valuable source of information about the relationship between pain and depression, which is incompletely understood. Subsequently, we apply our network motif detection approach to group-specific samples of ECNs to find patterns of directed interactions that may be considered as a characteristic of the group of patients or the group of controls, respectively. These characteristic patterns may shed light on the basic neural activity which occurs during the processing of painful stimuli in patients with major depression and in the healthy controls.

Due to the nature of the underlying EEG experiment, eight samples of ECNs may be considered. They are defined by all combinations of the group assignment (MD—patients suffering from major depression versus HC—healthy control subjects), the stimulated side (left versus right), as well as the time window with respect to the stimulus conditions (pre—time window before noxious stimulation versus post—time window directly following the stimulation, i.e., including the processing of the noxious stimulus). The nomenclature is MD-pre-left, MD-pre-right, MD-post-left, MD-post-right, HC-pre-left, and so forth. The sample size for the MD-post-right sample is fifteen, where the sample size equals sixteen for all other samples.

As directed graphs, effective connectivity networks consist of a nonempty finite set \mathcal{V} of vertices and a finite set \mathcal{E} of ordered pairs of distinct vertices called arcs or edges. An ordered pair (u_i, u_j) is called directed edge if it leaves vertex u_i and enters vertex u_j . It is denoted by $u_i \rightarrow u_j$ and u_i is called the tail and u_j is called the head of the edge. An ECN is represented by its adjacency matrix \mathcal{A} of size 9×9 where $\mathcal{A}_{ij} = 1$ if and only if the ECN contains the directed edge $u_i \rightarrow u_j$. Accordingly, a mutual edge is indicated by two entries in the adjacency matrix $\mathcal{A}_{ij} = 1$ and $\mathcal{A}_{ji} = 1$ and is denoted by $u_i \leftrightarrow u_j$. The effective connectivity networks are built by abstracting EEG-electrodes as vertices and modeling associated directed interactions by directed edges between

those vertices. The ECNs of the present study are small networks, each consisting of the same set of nine vertices that are pairwise differently labeled by associated EEG-electrode identifiers. For our approach, it is crucial that, due to the vertex labeling, all vertices are different. Most ECNs exhibit dense and intricate patterns of directed interactions. The mean number of edges in an ECN is 36.79 out of 72 possible edges. Moreover, ECNs do not contain multiple edges (edges with the same tail and the same head) and loops (edges whose tail and head coincide). Due to the properties of ECNs, their adjacency matrices are asymmetrical with 0 entries on the main diagonal. With two exceptions, all ECNs are connected networks. Examples of ECNs are depicted in Figure 1.

2.4. Network Motif Detection in a Sample of Directed Networks with Pairwise Different Vertex Labels. Dealing with network samples of directed networks with identical pairwise different vertex labeling instead of single networks without such labeling imposes certain constraints on the approach to network motif detection and also on the definition of a network motif. The most important constraint is that each subnetwork can occur at most only once in a single network, which affects the statistical analysis of subnetwork occurrences. It is not possible to assign significance to subnetwork counts in one network or in very small samples of networks. Therefore, motif detection in a sufficiently large sample of networks constitutes not only a novel approach to reveal common topological characteristics of all sample elements but is also a necessity. Given the pairwise different vertex labeling, two subnetworks are identical if and only if they share the same set of edges, that is, they have identical adjacency matrices. Therefore, isomorphic subnetworks do not exist and consequently it is unnecessary to address the problem of determining graph isomorphism for subnetworks. It is completely different for networks without vertex labeling. In the unlabeled case, different topological equivalence classes of subnetworks exist, also called motif classes or identities and each of them might consist of isomorphic subnetworks. For example, there are 13 equivalence classes of 3-subnetworks without vertex labeling comprising a total of 54 isomorphic subnetworks [18, 25]. In contrast, if the 3-subnetworks had pairwise different vertex labels, there are 54 different such subnetworks, each corresponding to one of the isomorphic subnetworks of the unlabeled case.

In order to keep the constraints given by the vertex labeling, one has to extend the original notion of network motifs [18, 19] to define the special case of network motifs of a sample of directed networks with pairwise different vertex labeling. Therefore, we define network motifs as small connected subnetworks which differ in their set of edges, as opposed to differing in their patterns of interconnections only, which appear in their sample of networks significantly more often than in random networks according to a suitable random graph model. In this way, we take the vertex labeling into account that does not only give us an advantage with respect to the computational complexity of our task to detect motifs but also has the important advantage

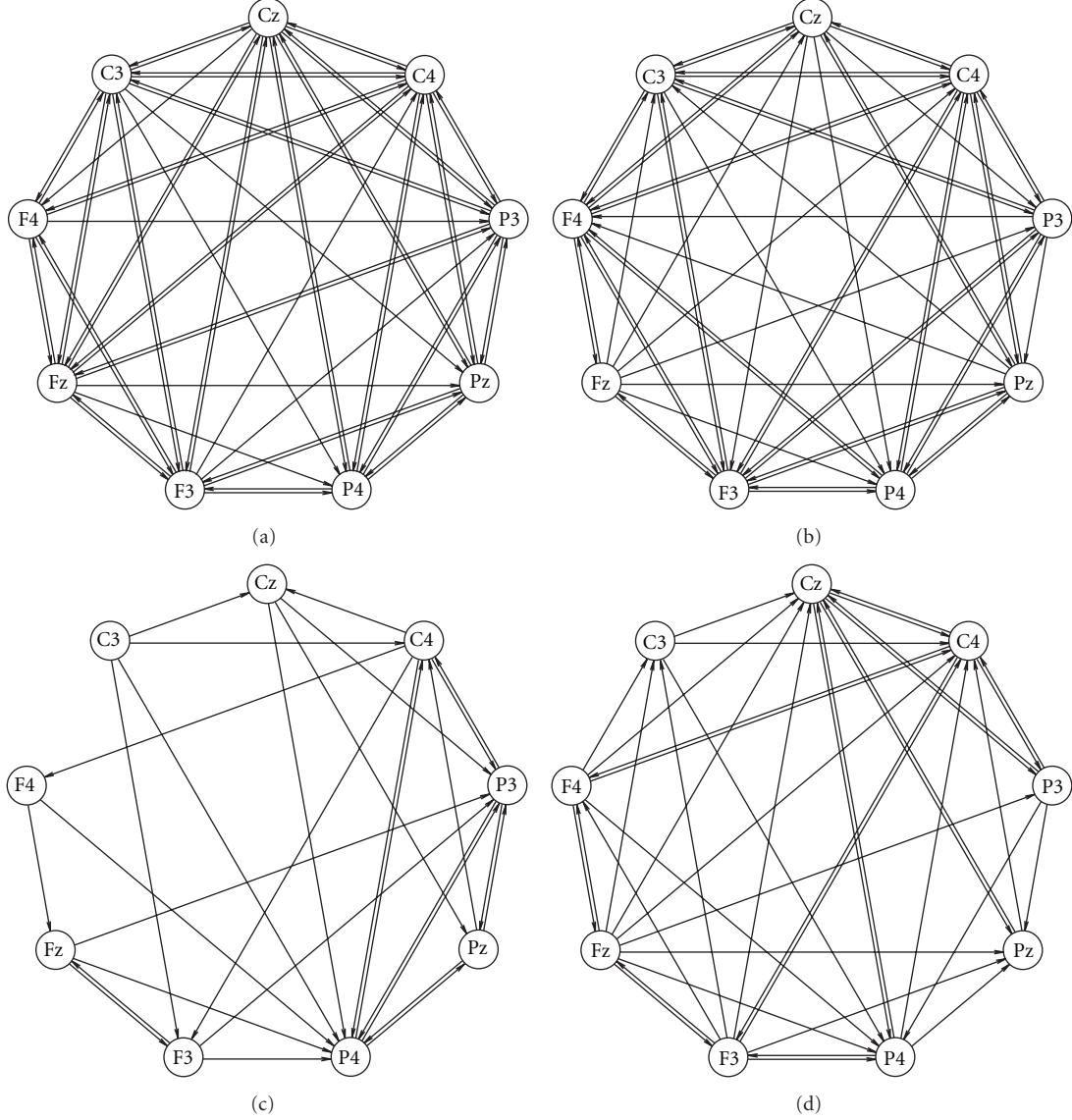


FIGURE 1: Examples of effective connectivity networks (ECNs). ECNs consist of nine vertices corresponding to EEG-electrodes that are pairwise differently labeled by associated EEG-electrode identifiers and directed edges corresponding to directed interactions between EEG-electrodes as indicated by significant gPDC values. In their wiring patterns, samples of ECNs contain intrinsic information about the processing of painful electrical stimuli in a group of patients suffering from major depression and a group of healthy controls.

of conserving the positional information of motifs in the network. This positional information is somehow associated with underlying neural processes and, therefore, is important for a subsequent functional interpretation of the results. If vertex labeling is discarded, then the only information obtained from motifs is about significant patterns of directed influences between EEG-electrodes. However, their localization would be missing which makes it unfeasible to functionally compare different instances of a motif in sets of vertex labeled networks.

2.4.1. Exhaustive Enumeration of Subnetworks. Let $\mathcal{N} = (\mathcal{N}_1, \dots, \mathcal{N}_n)$ be a sample of vertex-labeled directed networks $\mathcal{N}_i = (\mathcal{V}, \mathcal{E}_i)$ all having the same set \mathcal{V} of v vertices and a

particular set \mathcal{E}_i of directed edges. \mathcal{A}^i denotes the adjacency matrix that represents network \mathcal{N}_i . The first step in our approach is to explicitly enumerate all subnetworks of a certain size $v_S \geq 2$ in every network \mathcal{N}_i which is feasible due to the size of the networks and the sample size. Thereby, for each member network \mathcal{N}_i , every combination of v_S vertices is investigated with respect to the subnetwork induced by it. Subsequently, the number of occurrences of each induced subnetwork over the entire sample is counted. Based on these subnetwork counts, we analytically assign significance to subnetworks.

2.4.2. Testing Significant Subnetworks Occurrences. In order to identify subnetworks that occur significantly more often

than expected in random networks, a suitable model for such random networks is required. Such a model is called null model. Due to pairwise different vertex labels, each subnetwork can occur at most once in a network. Thus, the usual z -score approach [25, 26] cannot be applied. However, a suitable null model for labeled networks may be derived, if a sufficiently large sample of networks is available.

Let $0 \leq k_i \leq \nu(\nu - 1)$ be the number of edges of \mathcal{N}_i , and let

$$q = \frac{1}{n\nu(\nu - 1)} \sum_{i=1}^n k_i \quad (1)$$

be the normalized mean number of edges of the sample \mathcal{N} . Then, the i.i.d. variables \mathcal{A}_{kl}^0 , $1 \leq k \neq l \leq \nu$, with

$$\begin{aligned} P(\mathcal{A}_{kl}^0 = 1) &= q, \\ P(\mathcal{A}_{kl}^0 = 0) &= 1 - q, \end{aligned} \quad (2)$$

describe a random network $\mathcal{N}^0 = (\mathcal{V}, \mathcal{E}^0)$ with a mean number of edges $q\nu(\nu - 1)$. It provides the basis of the null model. Let S be an arbitrary subnetwork with at least $\nu_S \geq 2$ vertices of the set \mathcal{V} and η_S edges. Obviously, the subnetwork S can exhibit at most $\eta_{S_{\max}} = \nu_S(\nu_S - 1)$ edges. We are interested in the count that S occurs in the sample \mathcal{N} as subnetwork. For it, we define n i.i.d. random variables X_i by

$$X_i = \begin{cases} 1, & \text{if } S \text{ is a subnetwork of } \mathcal{N}^0 \\ 0, & \text{if } S \text{ is not a subnetwork of } \mathcal{N}^0. \end{cases} \quad (3)$$

Assuming the null model, the probability that S occurs as a subnetwork of \mathcal{N}^0 is by definition equal to

$$P(X_i = 1) = q^{\eta_S} \cdot (1 - q)^{\eta_{S_{\max}} - \eta_S} \quad (4)$$

for all $i = 1, \dots, n$. Since all sample networks \mathcal{N}_i are associated with the same null model, the count that S occurs in the sample \mathcal{N} as subnetwork is binomially distributed under the null model.

$$\sum_{i=1}^n X_i \sim B(n, q^{\eta_S} \cdot (1 - q)^{\eta_{S_{\max}} - \eta_S}). \quad (5)$$

Finally, all subnetworks of a certain size ν_S are tested with respect to a significant overrepresentation in the sample. Thus, an alpha-adjustment has to be applied. In the present study, generally the Bonferroni-Holm correction [55] with a multiple significance level of $\alpha = 0.05$ was adopted for all multiple test procedures to conservatively control the familywise error rate for all hypotheses at α in the strong sense instead of controlling the expected proportion of incorrectly rejected null hypotheses (false discovery rate).

3. Results and Discussion

We applied our approach to detect network motifs in eight group-specific samples of ECNs that were obtained from our effective connectivity data [51]. As a result of dismissing

interactions that are by definition less important, we reduce the information of the intricate patterns of directed interconnections of a sample of ECNs. We interpreted network motifs as patterns of characteristic interactions in a sample of ECNs. Because of the spatial information associated with the vertex labels, it makes sense to look even for 2-motifs in order to find significant interactions between two areas covered by the EEG-scheme. Furthermore, we were interested in characteristic interaction patterns that are represented by 3-motifs. We did not aim to detect motifs of a larger size because physiological interpretation of 2-motifs and 3-motifs is already difficult. Hence, detecting larger motifs does not seem to contribute much to the qualitative knowledge about effective connectivity networks. However, from a theoretical point of view, the detection of larger motifs is straightforward given that sufficiently large samples are available. Due to their small number, all 2-motifs detected by our approach could be presented in Table 1. In contrast, due to their large number, only those interesting 3-motifs whose occurrence is sample-specific or which occur in most samples of ECNs are presented in Table 2.

Table 1 illustrates that some 2-motifs represent functional connections that are present in the pre- as well as in the post-stimulus period, both for MD and HC. Typical examples are P4 \rightarrow C3 or Fz \leftrightarrow F4. Such connections might represent parts of the background activity or attentional processes which are independent of either group (MD, HC), time period (pre, post), or site of stimulation (left, right). Other motifs, for example, F3 \leftrightarrow Fz, are primarily present during the pre-stimulus period. Such motifs might represent processes of focusing attention to the next stimulus, preparation of the central resources, and so forth. Interestingly, there are several motifs that are specific to MD patients only, for example, C4 \leftrightarrow F4, Cz \leftrightarrow C4, or Cz \leftrightarrow Pz, while others are specific to HC subjects, for example, C3 \leftrightarrow Cz. The motifs specific to MD patients occur more often during the pre-stimulus period. They are concentrated on the central electrodes and electrodes on the right hemisphere. This might reflect the role of the right hemisphere in the processing of emotions and mood, especially in MD patients [45, 47]. In contrast, the motif specific to HC subjects is the only one that is also specific for the processing in the post-stimulus period. Therefore, it is probable that it represents the processing of the noxious stimulus itself. One might wonder that MD patients do not exhibit such a motif (or any other motif specific during the post-stimulus period); however, it should be mentioned that MD patients have been found to exhibit higher pain thresholds [42, 46], lower sensitivity to experimental nociceptive stimulation [46, 56], and/or lower processing of C-fiber nociceptive activation [57].

Motifs of size 3 (Table 2) also show differences concerning groups, time period, and stimulated site. Thus, the motif P4-Fz-F3 (motif 1, Table 2) is present in the pre- and the post-stimulus period for left and right stimulation both for MD and HC, probably representing baseline activity or brain activity that is independent from stimulation and group. Other motifs can be found only during the pre-stimulus period, for example, P4-Cz-F3 (motif 2, Table 2) or

TABLE 1: The mean number of all 2-motifs in the eight samples of effective connectivity networks (ECNs). The 2-motifs represent important interactions before and during the processing of painful electrical stimuli. The samples originate from all combinations of the group assignment. MD—patients suffering from major depression, HC—healthy control subjects, left and right—stimulated side, pre and post—time window with respect to the stimulus condition.

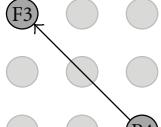
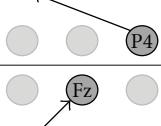
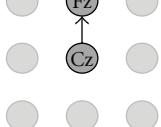
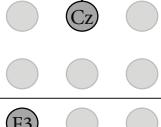
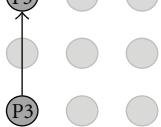
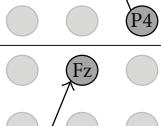
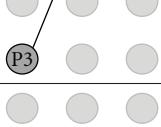
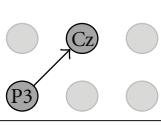
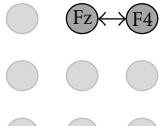
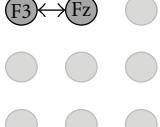
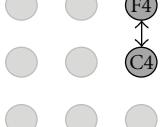
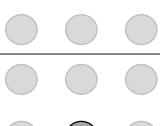
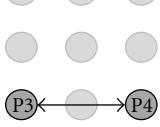
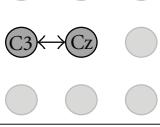
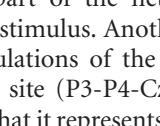
Motif	MD				HC			
	left		right		left		right	
	pre	post	pre	post	pre	post	pre	post
(1) 	13.0	11.0	13.0	—	—	—	—	—
(2) 	11.0	11.0	14.0	11.0	12.0	13.0	11.0	12.0
(3) 	—	—	11.0	—	—	—	—	—
(4) 	—	—	13.0	—	11.9	—	—	11.0
(5) 	—	—	11.0	—	—	—	—	—
(6) 	—	11.0	11.0	—	—	—	—	11.0
(7) 	—	—	—	—	11.0	—	—	—
(8) 	—	—	—	—	—	—	11.0	—
(9) 	12.2	12.8	12.2	11.2	11.0	—	13.0	14.0

TABLE 1: Continued.

Motif	MD				HC			
	left		right		left		right	
	pre	post	pre	post	pre	post	pre	post
(10) 	—	—	12.0	—	11.0	—	10.0	—
(11) 	—	—	12.0	—	—	—	—	—
(12) 	—	—	11.3	—	—	—	—	—
(13) 	—	—	11.9	—	—	—	—	—
(14) 	—	14.0	11.0	—	—	12.0	—	—
(15) 	—	—	—	—	—	13.3	—	13.2

Cz-Fz-C3 (motif 4, Table 2). It is likely that these motifs are part of the network that prepares the brain for the next stimulus. Another motif occurred only before and after stimulations of the left site and before stimulation of the right site (P3-P4-Cz, motif 7, Table 2). One might speculate that it represents attentional processes before stimulation is also involved in the information processing when the left hand was stimulated. Many of these motifs involve the right parietal electrode P4 further supporting the notion of possible attentional processes. There are also motifs specifically found for stimulations of the right hand (C4-F4-Fz, motif 9 and P4-C3-Fz, motif 10, both in Table 2). All the motifs of size 3 mentioned above are independent of the group, thus representing activity for both HC and MD subjects. Interestingly, there are also motifs that differentiate between MD patients and HC subjects. Thus, the P3-P4-Fz (motif 5, Table 2) occurred in MD patients during the pre-stimulation period whereas it was found only after stimulation in HC. This might be a hint that the processing during

TABLE 2: The mean number of 3-motifs with interesting similarities and differences in the eight samples of effective connectivity networks (ECNs). The 3-motifs represent important patterns of interactions before and during the processing of painful electrical stimuli. The samples originate from all combinations of the group assignment. MD—patients suffering from major depression, HC—healthy control subjects, left and right—stimulated side, pre and post—time window with respect to the stimulus condition.

Motif	MD				HC			
	left		right		left		right	
	pre	post	pre	post	pre	post	pre	post
(1)		4.0	7.0	9.0	5.0	7.0	—	5.0 8.0
(2)		5.0	—	4.0	—	5.0	—	—
(3)		4.5	5.0	7.0	4.0	6.0	—	— 5.0
(4)		4.2	—	6.3	—	—	—	4.2 —
(5)		4.5	6.0	6.0	6.0	—	—	— 4.0
(6)		4.0	—	4.3	—	5.0	—	4.2 —
(7)		5.5	4.5	4.0	—	4.4	5.0	5.9 —
(8)		—	—	7.0	—	7.0	—	5.0 —
(9)		—	—	5.0	5.0	—	—	3.0 5.0

TABLE 2: Continued.

Motif	MD				HC			
	left		right		left		right	
	pre	post	pre	post	pre	post	pre	post
(10)		—	—	7.0	6.0	4.0	—	4.0 5.0
(11)		—	4.5	8.5	7.0	5.6	—	5.5 6.0
(12)		—	—	5.0	6.0	4.0	—	4.0 4.0
(13)		—	—	—	—	4.0	—	3.0 5.0
(14)		—	—	—	4.0	—	4.1	— 5.0
(15)		—	—	4.3	—	4.3	—	4.0 —
(16)		—	—	4.5	7.3	—	—	— 4.3

the prestimulation time in MD involves some networks that resemble the (possibly affective) processing of noxious stimulation in HC subjects.

These results offer a number of intriguing insights into various patterns of directed interactions associated with the processing of painful and, therefore, salient stimuli, characteristic of both groups over the course of time during the experiment. However, the concept of motif detection remains controversial and questions remain. First, it is clear that motif detection misses any functional meaningful subnetworks that appear only infrequently. Conversely, subnetworks that appear with significant frequency are not necessarily important for the functioning of their network. Another criticism refers to the claim that the occurrence of specific motifs is characteristic for a certain network or

a type of networks. It seems that some motifs of a network's motif distribution might occur due to contingencies in the network structure and due to topological effects known as spatial clustering (closeness of vertices in topological space or in attribute space) [31]. A test for an underlying geometric arrangement in real-world network topology has been proposed in [58]. The test is basically a comparison of invariant ratios of the numbers of certain subnetworks in geometric random network models with the same ratios obtained from the real-world network. In the same study, it has been found that the ratios in exemplary real-world networks generally differed from the ratios in geometric random networks. Thus, the authors concluded that network motifs in many real-world networks are not solely captured by geometric constraints but instead arise due to additional functional optimization of network topology. Likewise, a preceding study used subnetwork significance profiles and subnetwork ratios obtained from exemplary real-world networks and either geometric- or preferential-attachment networks to show that spatial clustering does not affect the number of occurrences of the majority of subnetworks and can also be ruled out as the primary mechanism that forms the structure of the real-world networks [59].

The potential ambiguity of the structure-function relationship of subnetworks and the influence of selection pressure versus variability operators on network topology, as well as the role of entanglement of subnetworks with the rest of the network, has also been debated [60]. In the context of assigning functionality to motifs, it has been argued that topological information on subnetworks must be complemented with information on parameters which describe the dynamic properties of the system, as motifs show different (and even opposing) dynamic behavior for different ranges of parameter values [61]. It has also been shown that an alleged dynamic behavior of motifs is strongly affected by the global and local dynamics of the entire network since motifs are not isolated within the network but rather are functionally interacting with many other surrounding parts of the network [61]. Investigation of the functional dependence of motifs on their context and the incorporation of parameters in the assignment of functionality to motifs is lacking in current studies; answering these analytical challenges remains a topic for further research. However, isolated network motifs have been tested experimentally for their regulatory functions as recurring circuits in bacteria and yeast transcription networks [62, 63]. The experimental studies confirmed theoretical predictions and could assign specific modes of molecular information processing to distinct motifs in these networks. Therefore, it has been shown that network motifs appear to be main building blocks of transcription networks. In principle, the role of network motifs in different systems can be examined experimentally, too.

Yet although these critiques underline potential limitations and pitfalls in assigning functionality to motifs (which is the reason for detecting them), it surely does not invalidate the concept of furthering the understanding of a network's functionality and uncovering its design principles by first analyzing local functional substructures, and then combining this information to infer network behavior at a global

level. Moreover, this criticism in its entirety does not hold for our use and interpretation of network motifs, because we are primarily interested in obtaining patterns of interactions that are overrepresented in a sample of networks. It is solely this overrepresentation that allows for an interpretation of these network motifs as a characteristic of this sample. At the same time, we ignore those patterns of interactions that are not overrepresented. In this respect, our approach might be seen as a tool that simplifies the intricate topology of each member in a sample of networks by thinning out interactions that are less important for the sample of networks. Finally, after this simplification, we are able to compare different samples of networks, for example, samples of ECNs. Currently, given the outlined criticisms and lack of neurophysiological knowledge on pain processing, an understanding of the information processing roles network motifs carry out in ECNs is not yet attainable.

We have applied network motif detection to unipolar data with a linked-ears reference. It has been shown previously that the reference might affect the results of such analyses. Specifically for coherence estimates, it is not possible to accurately predict reference effects without an accurate volume conductor model and prior knowledge of all source locations [64]. Thus, the current underlying connectivity analysis applies to the sensor space with linked ears as reference rather than to the source space. Consequently, the motif detection focuses on network motifs at the sensor level. Therefore, the current view on anatomical locations of motifs might only serve as a cautious hint with reference to anatomical sources.

The underlying gPDC analysis has been performed on the basis of SEPs, where a multitrial estimator was applied to estimate the autoregressive model parameters [51]. For it, all raw single trials were provided separately to the estimator without any prior averaging. Thus, an explicit separation of ongoing and evoked activity [65] was not carried out. In the post-stimulus condition, the identified effective connectivity patterns, as well as the derived network motifs contain a certain amount of information associated to ongoing activity. For this reason a pre-stimulus condition was also studied in order to investigate effective interactions based solely on ongoing activity. It has been shown in [51] that the stimulus resulted in significant gPDC changes in both groups. As a consequence, we show alterations of network motif appearances associated to the stimulus.

The design of a suitable null model defines the notion of randomness and is crucial for distinguishing regular topological effects from true topological contingencies in the sample of ECNs and thus is crucial for obtaining valid results [31]. At the present time, there is no established theoretical background for choosing null models that fit to given network data and thus it is not clear which network properties might be incorporated into a good null model. The null model widely employed in motif detection preserves the degree sequence of the input network, which is a basic property on the vertex level that ultimately affects many other properties of the network. Studies that make use of this somewhat more elaborate null model rely on algorithms for generating very large sets of random networks out of

the original network that all hold the desired property. These simulations are very time consuming and so is the following counting of subnetwork appearances in the obtained random ensemble. On the other hand, our analytical statistical test is computed much faster (within seconds implemented in the MATLAB programming language) than simulating random networks and counting their subnetwork occurrences, but at the cost of simpler assumptions being made for the null model which accounts for the mean number of edges of the input network sample.

4. Conclusions

We have presented an approach to analytically detect network motifs in a sample of directed networks with pairwise different vertex labels. The importance of choosing an appropriate null model random network that contrasts topological regularities of the input networks with topological contingency is outlined. Clearly, a refinement of our analytical null model, which accounts for the average number of edges in a sample of networks, is desirable. Such a refinement is currently under investigation. Nonetheless, we have demonstrated that our approach to network motif detection is suitable to act as a filter to reveal locatable patterns of directed interactions that might be interpreted as characteristic for each of several group-specific samples of ECNs. These networks originate from effective connectivity data obtained in our previous study that investigated cortical activity before and after painful stimulation of patients with major depression and healthy control subjects [51]. The detected motifs on the one hand yield a compact description of recurring important topological elements in a sample of ECNs. On the other hand, they allow for a comparison of different samples of ECNs, which was as yet not attainable. The sample-specific network motifs of ECNs can now be investigated in more depth to gain further understanding of neurophysiological processes in both groups during the anticipation and processing of painful stimuli. This in turn should contribute to a deeper understanding of the relationship between pain and depression.

Acknowledgments

This work was supported by the Grants WI 1166/9-2 and NI 369/7 of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). The authors thank Elizabeth Ahrens-Kley for grammatical aid.

References

- [1] B. Vogelstein, D. Lane, and A. J. Levine, "Surfing the p53 network," *Nature*, vol. 408, no. 6810, pp. 307–310, 2000.
- [2] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging discipline," *Molecular Systems Biology*, vol. 2, Article ID 2006.0028, pp. 1–14, 2006.
- [3] R. McDaniel and R. Weiss, "Advances in synthetic biology: on the path from prototypes to applications," *Current Opinion in Biotechnology*, vol. 16, no. 4, pp. 476–483, 2005.
- [4] M. J. Keeling, L. Danon, A. P. Ford et al., "Networks and the epidemiology of infectious disease," *Interdisciplinary Perspectives on Infectious Diseases*, vol. 2011, Article ID 284909, 28 pages, 2011.
- [5] C. J. Stam and J. C. Reijneveld, "Graph theoretical analysis of complex networks in the brain," *Nonlinear Biomedical Physics*, vol. 1, article no. 3, 2007.
- [6] D. S. Bassett and E. Bullmore, "Small-world brain networks," *Neuroscientist*, vol. 12, no. 6, pp. 512–523, 2006.
- [7] Y. He, Z. J. Chen, and A. C. Evans, "Small-world anatomical networks in the human brain revealed by cortical thickness from MRI," *Cerebral Cortex*, vol. 17, no. 10, pp. 2407–2419, 2007.
- [8] C. J. Stam, B. F. Jones, G. Nolte, M. Breakspear, and P. Scheltens, "Small-world networks and functional connectivity in Alzheimer's disease," *Cerebral Cortex*, vol. 17, no. 1, pp. 92–99, 2007.
- [9] Y. Liu, M. Liang, Y. Zhou et al., "Disrupted small-world networks in schizophrenia," *Brain*, vol. 131, no. 4, pp. 945–961, 2008.
- [10] S. Micheloyannis, E. Pachou, C. J. Stam et al., "Small-world networks and disturbed functional connectivity in schizophrenia," *Schizophrenia Research*, vol. 87, no. 1-3, pp. 60–66, 2006.
- [11] B. C. Bernhardt, Z. Chen, Y. He, A. C. Evans, and N. Bernasconi, "Graph-theoretical analysis reveals disrupted small-world organization of cortical thickness correlation networks in temporal lobe epilepsy," *Cerebral Cortex*, vol. 21, no. 9, pp. 2147–2157, 2011.
- [12] M. Catani and M. Mesulam, "What is a disconnection syndrome?" *Cortex*, vol. 44, no. 8, pp. 911–913, 2008.
- [13] X. Delbeuck, M. Van Der Linden, and F. Collette, "Alzheimer's disease as a disconnection syndrome?" *Neuropsychology Review*, vol. 13, no. 2, pp. 79–92, 2003.
- [14] K. J. Friston, "Schizophrenia and the disconnection hypothesis," *Acta Psychiatrica Scandinavica*, vol. 99, no. 395, pp. 68–79, 1999.
- [15] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [16] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [17] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [18] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.
- [19] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of Escherichia coli," *Nature Genetics*, vol. 31, no. 1, pp. 64–68, 2002.
- [20] T. I. Lee, N. J. Rinaldi, F. Robert et al., "Transcriptional regulatory networks in *Saccharomyces cerevisiae*," *Science*, vol. 298, no. 5594, pp. 799–804, 2002.
- [21] R. Milo, S. Itzkovitz, N. Kashtan et al., "Superfamilies of Evolved and Designed Networks," *Science*, vol. 303, no. 5663, pp. 1538–1542, 2004.
- [22] R. J. Prill, P. A. Iglesias, and A. Levchenko, "Dynamic properties of network motifs contribute to biological network organization," *PLoS Biology*, vol. 3, no. 11, pp. 1881–1892, 2005.
- [23] I. Albert and R. Albert, "Conserved network motifs allow protein-protein interaction prediction," *Bioinformatics*, vol. 20, no. 18, pp. 3346–3352, 2004.
- [24] E. Yeger-Lotem, S. Sattath, N. Kashtan et al., "Network motifs in integrated cellular networks of transcription-regulation

- and protein-protein interaction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 16, pp. 5934–5939, 2004.
- [25] O. Sporns and R. Kötter, "Motifs in Brain Networks," *PLoS Biology*, vol. 2, no. 11, 2004.
- [26] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon, "Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs," *Bioinformatics*, vol. 20, no. 11, pp. 1746–1758, 2004.
- [27] S. Wernicke, "A faster algorithm for detecting network motifs," in *Proceedings of the 5th Workshop on Algorithms in Bioinformatics (WABI '05)*, vol. 3692, pp. 165–177, 2005.
- [28] G. Ciriello and C. Guerra, "A review on models and algorithms for motif discovery in protein-protein interaction networks," *Briefings in Functional Genomics and Proteomics*, vol. 7, no. 2, pp. 147–156, 2008.
- [29] F. Schreiber and H. Schwobermeyer, "Frequency concepts and pattern detection for the analysis of motifs in networks," *Transactions on Computational Systems Biology III*, vol. 3737, pp. 89–104, 2005.
- [30] P. Foggia, C. Sansone, and M. Vento, "A performance comparison of five algorithms for graph isomorphism," in *Proceedings of the 3rd IAPR TC-15 Workshop on Graph-based Representations in Pattern Recognition*, pp. 188–199, 2001.
- [31] Y. Artzy-Randrup, S. J. Fleishman, N. Ben-Tal, and L. Stone, "Comment on 'Network motifs: simple building blocks of complex networks' and 'Superfamilies of evolved and designed networks,'" *Science*, vol. 305, no. 5687, p. 1107, 2004.
- [32] O. D. King, S. Itzkovitz, R. Milo, N. Kashtan, M. E. J. Newman, and U. Alon, "Comment on 'subgraphs in random networks,'" *Physical Review E*, vol. 70, no. 5, Article ID 058101, pp. 1–58102, 2004.
- [33] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon, "On the uniform generation of random graphs with prescribed degree sequences," Arxiv preprint condmat/ 0312028, 2003.
- [34] M. E. J. Newman, "Random graphs as models of networks," Arxiv preprint condmat/ 0202208, 2002.
- [35] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, no. 2, Article ID 026118, pp. 261181–261187, 2001.
- [36] R. Kannan, P. Tetali, and S. Vempala, "Simple Markov-chain algorithms for generating bipartite graphs and tournaments," in *Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 193–200, Society for Industrial and Applied Mathematics, 1997.
- [37] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, 2002.
- [38] E. Ziv, R. Koytcheff, M. Middendorf, and C. Wiggins, "Systematic identification of statistically significant network measures," *Physical Review E*, vol. 71, no. 1, Article ID 16110, 2005.
- [39] L. Astolfi, F. Cincotti, D. Mattia et al., "Time-varying cortical connectivity estimation from noninvasive, high-resolution EEG recordings," *Journal of Psychophysiology*, vol. 24, no. 2, pp. 83–90, 2010.
- [40] C. Ligges, M. Ungureanu, M. Ligges, B. Blanz, and H. Witte, "Understanding the time variant connectivity of the language network in developmental dyslexia: new insights using Granger causality," *Journal of Neural Transmission*, vol. 117, no. 4, pp. 529–543, 2010.
- [41] T. Milde, L. Leistritz, L. Astolfi et al., "A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials," *NeuroImage*, vol. 50, no. 3, pp. 960–969, 2010.
- [42] S. Lautenbacher, J. Sernal, W. Schreiber, and J. C. Krieg, "Relationship between clinical pain complaints and pain sensitivity in patients with depression and panic disorder," *Psychosomatic Medicine*, vol. 61, no. 6, pp. 822–827, 1999.
- [43] L. R. Miller and A. Cano, "Comorbid chronic pain and depression: who is at risk?" *Journal of Pain*, vol. 10, no. 6, pp. 619–627, 2009.
- [44] S. Klauenberg, C. Maier, H. J. Assion et al., "Depression and changed pain perception: hints for a central disinhibition mechanism," *Pain*, vol. 140, no. 2, pp. 332–343, 2008.
- [45] I. A. Strigo, A. N. Simmons, S. C. Matthews, A. D. Craig, and M. P. Paulus, "Association of major depressive disorder with altered functional brain response during anticipation and processing of heat pain," *Archives of General Psychiatry*, vol. 65, no. 11, pp. 1275–1284, 2008.
- [46] K. J. Bär, S. Brehm, M. K. Boettger, S. Boettger, G. Wagner, and H. Sauer, "Pain perception in major depression depends on pain modality," *Pain*, vol. 117, no. 1-2, pp. 97–103, 2005.
- [47] K. J. Bär, G. Wagner, M. Koschke et al., "Increased Prefrontal Activation During Pain Perception in Major Depression," *Biological Psychiatry*, vol. 62, no. 11, pp. 1281–1287, 2007.
- [48] S. Lautenbacher and J. C. Krieg, "Pain perception in psychiatric disorders: a review of the literature," *Journal of Psychiatric Research*, vol. 28, no. 2, pp. 109–122, 1994.
- [49] J. Terhaar, F. C. Viola, M. Franz, S. Berger, K.-J. Bär, and T. Weiss, "Differential processing of laser stimuli by A δ and C fibres in major depression," *Pain*, vol. 152, no. 8, pp. 1796–1802, 2011.
- [50] G. D. Iannetti and A. Mouraux, "From the neuromatrix to the pain matrix (and back)," *Experimental Brain Research*, vol. 205, no. 1, pp. 1–12, 2010.
- [51] L. Leistritz, T. Weiss, J. Ionov, K. J. Bär, W. H. R. Miltner, and H. Witte, "Connectivity analysis of somatosensory evoked potentials to noxious intracutaneous stimuli in patients with major depression," *Methods of Information in Medicine*, vol. 49, no. 5, pp. 484–491, 2010.
- [52] W. Meissner, T. Weiss, R. H. Trippe, H. Hecht, C. Krapp, and W. H. R. Miltner, "Acupuncture Decreases Somatosensory Evoked Potential Amplitudes to Noxious Stimuli in Anesthetized Volunteers," *Anesthesia and Analgesia*, vol. 98, no. 1, pp. 141–147, 2004.
- [53] T. Weiss, K. Kumpf, J. Ehrhardt, I. Gutberlet, and W. H. R. Miltner, "A bioadaptive approach for experimental pain research in humans using laser-evoked brain potentials," *Neuroscience Letters*, vol. 227, no. 2, pp. 95–98, 1997.
- [54] L. A. Baccala, K. Sameshima, and D. Y. Takahashi, "Generalized partial directed coherence," in *Proceedings of the 15th International Conference on Digital Signal Processing (DSP '07)*, pp. 163–166, IEEE, July 2007.
- [55] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.
- [56] K. J. Bär, S. Brehm, M. K. Boettger, G. Wagner, S. Boettger, and H. Sauer, "Decreased sensitivity to experimental pain in adjustment disorder," *European Journal of Pain*, vol. 10, no. 5, pp. 467–471, 2006.
- [57] T. Weiss, D. A. Giersch, H. Sauer, W. H. R. Miltner, and K.-J. Bär, "Perception to laser heat stimuli in depressed patients is reduced to A δ - and selective C-fiber stimulation," *Neuroscience Letters*, vol. 498, no. 1, pp. 89–92, 2011.

- [58] S. Itzkovitz and U. Alon, "Subgraphs and network motifs in geometric networks," *Physical Review E*, vol. 71, no. 2, Article ID 026117, pp. 026117/1–026117/9, 2005.
- [59] Y. Artzy-Randrup, S. J. Fleishman, N. Ben-Tal, and L. Stone, "Response to comment on 'network motifs: simple building blocks of complex networks' and 'Superfamilies of evolved and designed networks,'" *Science*, vol. 305, no. 5687, p. 1107, 2004.
- [60] J. F. Knabe, C. L. Nehaniv, and M. J. Schilstra, "Do motifs reflect evolved function? No convergent evolution of genetic regulatory network subgraph topologies," *BioSystems*, vol. 94, no. 1-2, pp. 68–74, 2008.
- [61] P. J. Ingram, M. P. H. Stumpf, and J. Stark, "Network motifs: structure does not determine function," *BMC Genomics*, vol. 7, p. 108, 2006.
- [62] U. Alon, "Network motifs: theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, no. 6, pp. 450–461, 2007.
- [63] O. Shoval and U. Alon, "SnapShot: network Motifs," *Cell*, vol. 143, no. 2, pp. 326–e1, 2010.
- [64] P. L. Nunez, R. Srinivasan, A. F. Westdorp et al., "EEG coherence I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 5, pp. 499–515, 1997.
- [65] X. Wang, Y. Chen, and M. Ding, "Estimating Granger causality after stimulus onset: a cautionary note," *NeuroImage*, vol. 41, no. 3, pp. 767–776, 2008.

Research Article

Measuring Connectivity in Linear Multivariate Processes: Definitions, Interpretation, and Practical Analysis

Luca Faes, Silvia Erla, and Giandomenico Nollo

Laboratorio Biosegnali, Dipartimento di Fisica & BIOTech, Università di Trento, via delle Regole 101, 38123 Mattarello, Trento, Italy

Correspondence should be addressed to Luca Faes, luca.faes@unitn.it

Received 28 October 2011; Revised 22 February 2012; Accepted 3 March 2012

Academic Editor: Dimitris Kugiumtzis

Copyright © 2012 Luca Faes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This tutorial paper introduces a common framework for the evaluation of widely used frequency-domain measures of coupling (coherence, partial coherence) and causality (directed coherence, partial directed coherence) from the parametric representation of linear multivariate (MV) processes. After providing a comprehensive time-domain definition of the various forms of connectivity observed in MV processes, we particularize them to MV autoregressive (MVAR) processes and derive the corresponding frequency-domain measures. Then, we discuss the theoretical interpretation of these MVAR-based connectivity measures, showing that each of them reflects a specific time-domain connectivity definition and how this results in the description of peculiar aspects of the information transfer in MV processes. Furthermore, issues related to the practical utilization of these measures on real-time series are pointed out, including MVAR model estimation and significance assessment. Finally, limitations and pitfalls arising from model mis-specification are discussed, indicating possible solutions and providing practical recommendations for a safe computation of the connectivity measures. An example of estimation of the presented measures from multiple EEG signals recorded during a combined visuomotor task is also reported, showing how evaluation of coupling and causality in the frequency domain may help describing specific neurophysiological mechanisms.

1. Introduction

Multivariate (MV) time series analysis is nowadays extensively used to investigate the concept of connectivity in dynamical systems. Connectivity is evaluated from the joint description of multiple time series collected simultaneously from the considered system. Applications of this approach are ubiquitous in the analysis of experimental time series recorded in various research fields, ranging from economics to biomedical sciences. In neuroscience, the concept of brain connectivity [1] plays a central role both in the understanding of the neurophysiological mechanisms of interaction among different areas of the brain, and in the development of indexes for the assessment of mechanism impairment in pathological conditions (see, e.g., [2] and references therein). The general term “brain connectivity” encompasses different modes, each making reference to

specific aspects of how brain areas interact. In particular, “functional connectivity” refers to evaluation of statistical dependencies between spatially distributed neuronal units, while “effective connectivity” refers to the description of networks of directional effects of one neural unit over another [3]. In the context of time series analysis, the notions of functional and effective connectivity can be investigated, respectively, in terms of *coupling*, that is, the presence of interactions, and of *causality*, that is, the presence of driver-response relationships, between two neurophysiological time series taken from the available MV data set.

The assessment of coupling and causality in MV processes may be performed following either linear or nonlinear time series analysis approaches [2, 4]. While nonlinear methods are continuously under development [5–10] and offer the intriguing possibility of studying complex signal interactions, linear signal processing approaches [11] are

extensively used in MV neurophysiological time series analysis. The main reason for the popularity of linear methods lies in the fact that they are strictly related to the frequency-domain representation of multichannel data [12, 13], and thus, lend themselves to the representation of biological signals which are rich of oscillatory content. In physiological systems, the linear frequency-domain representation favors the characterization of connectivity between specific oscillatory components such as the EEG rhythms [14].

In the linear signal processing framework, connectivity is very often formalized in the context of an MV autoregressive (MVAR) representation of the available time series, which allows to derive time- and frequency-domain pictures, respectively, by the model coefficients and by their spectral representation. Accordingly, several frequency-domain measures of connectivity have been introduced and applied in recent years. Coupling is traditionally investigated by means of the coherence (Coh) and the partial coherence (PCoh), classically known, for example, from Kay [15] or Bendat and Piersol [16]. Measures able to quantify causality in the frequency-domain were proposed more recently, the most used being the directed transfer function (DTF) [12, 17], the directed coherence (DC) [18], and the partial directed coherence (PDC) [19], the latter repeatedly refined after its original formulation [20–22]. These measures are widely used for the analysis of interactions among physiological time series, and—in particular—to characterize brain connectivity [23–31]. Recent studies have proposed deeper interpretation of frequency-domain connectivity measures [21, 32], as well as comparison on both simulated and real physiological time series [11, 33]. Despite this large body of work, the interpretation of frequency-domain coupling and causality measures is not always straightforward, and this may lead to erroneous descriptions of connectivity and related mechanisms. Examples of ambiguities emerged in the interpretation of these measures are the debates about the ability of PCoh to measure some forms of causality [34, 35], about the specific kind of causality which is reflected by the DTF and DC measures [17, 19, 36], and about whether the PDC could be suitably re-normalized to make its modulus able to reflect meaningfully the strength of coupling [22, 32].

In order to settle these interpretability issues, a joint description of the different connectivity measures, as well as a contextualization in relation to well-defined time-domain concepts, is required. According to this need, the present paper has a tutorial character such that—instead of proposing new measures—it is aimed to enhance the interpretability and favour the utilization of existing frequency-domain connectivity measures based on MVAR modelling. To this end, we introduce a common framework for the evaluation of Coh, PCoh, DC/DTF, and PDC from the frequency-domain representation of MVAR processes, which is exploited to relate the various measures to each other as well as to the specific coupling or causality definition which they underlie. After providing a comprehensive definition of the various forms of connectivity observed in MV processes, we particularize them to MVAR processes and derive the corresponding frequency-domain measures. Then, we discuss the theoretical interpretation of these measures,

showing how they are able to describe peculiar aspects of the information transfer in MV time series. Further, we point out issues related to practical estimation, limitations, and recommendations for the utilization of these measures on real MV time series. An example of estimation of the presented measures from multiple EEG signals recorded during a sensorimotor integration experiment is finally presented to illustrate their practical applicability.

2. Connectivity Definitions in the Time Domain

2.1. Multivariate Closed-Loop Processes. Let us consider M stationary stochastic processes y_m , $m = 1, \dots, M$, collected in the multivariate (MV) vector process $\mathbf{Y} = [y_1, \dots, y_M]^T$. Without loss of generality, we assume that the processes are real-valued, defined at discrete time ($y_m = \{y_m(n)\}$; e.g., are sampled versions of the continuous time processes $y_m(t)$, taken at the times $t_n = nT$, with T the sampling period) and have zero mean ($E[y_m(n)] = 0$, where $E[\cdot]$ is the statistical expectation operator). An MV closed loop vector process of order p is defined expressing the present value of each scalar process, $y_m(n)$, as a function of the p past values of all processes, collected in $Y_l = \{y_l(n-1), \dots, y_l(n-p)\}$ ($l, m = 1, \dots, M$):

$$y_m(n) = f_m(Y_1, \dots, Y_M) + u_m(n), \quad (1)$$

where u_m are independent white noise processes describing the error in the representation. Note that the definition in (1) limits to past values only the possible influences of one process to another, excluding instantaneous effects (i.e., effects occurring within the same lag). The absence of instantaneous effects is denoted as strict causality of the closed loop MV process [37, 38] and will be assumed henceforth.

Given two processes y_i and y_j of the closed-loop, the general concept of *connectivity* can be particularized to the study of *causality* or *coupling* between y_i and y_j , which investigate, respectively, directional or non-directional properties of the considered pairwise interaction. With the aim of supporting interpretation of the frequency-domain connectivity measures presented in Section 3, we state now specific time-domain definitions of coupling and causality valid for an MV closed-loop process (see Table 1). *Direct causality* from y_j to y_i , $y_j \rightarrow y_i$, exists if the prediction of $y_i(n)$ based on $\{Y_1, \dots, Y_M\}$ is better (i.e., yields a lower prediction error) than the prediction of $y_i(n)$ based on $\{Y_1, \dots, Y_M\} \setminus Y_j$. *Causality* from y_j to y_i , $y_j \Rightarrow y_i$, exists if a cascade of direct causality relations $y_j \rightarrow y_m \cdots \rightarrow y_i$ occurs for at least one $m \in \{1, \dots, M\}$; if $m = i$ or $m = j$ causality reduces to direct causality, while for $m \neq i, m \neq j$, the causality relation is indirect. *Direct coupling* between y_i and y_j , $y_i \leftrightarrow y_j$, exists if $y_i \rightarrow y_j$ or $y_j \rightarrow y_i$. *Coupling* between y_i and y_j , $y_i \Leftrightarrow y_j$, exists if $y_i \Rightarrow y_j$ or $y_j \Rightarrow y_i$. The rationale of these connectivity definitions is grounded on the very popular notion of Granger causality, as originally introduced by the seminal paper of Granger for a bivariate closed loop linear stochastic process [39], and on intuitive generalizations aimed at moving from the study of causality

TABLE 1: Connectivity definitions and conditions for their existence. (See text for details).

Definition		MV closed-loop process	MVAR process, time domain	MVAR process, Frequency domain
Direct causality	$y_j \rightarrow y_i$	Knowledge of y_j improves prediction of $y_i(n)$	$a_{ij}(k) \neq 0$	$\pi_{ij}(f) \neq 0$
Causality	$y_j \Rightarrow y_i$	$y_j \rightarrow y_m \dots \rightarrow y_i$	$a_{mj}(k) \neq 0, \dots, a_{im}(k) \neq 0$	$\gamma_{ij}(f) \neq 0$
Direct coupling	$y_i \leftrightarrow y_j$	$y_i \rightarrow y_j$ or $y_j \rightarrow y_i$	$a_{ji}(k) \neq 0$ or $a_{ij}(k) \neq 0$	$\Pi_{ij}(f) \neq 0$
Spurious direct coupling		$y_i \rightarrow y_m$ and $y_j \rightarrow y_m$	$a_{mi}(k) \neq 0$ and $a_{mj}(k) \neq 0$	
Coupling	$y_i \Leftrightarrow y_j$	$y_i \Rightarrow y_j$ or $y_j \Rightarrow y_i$	$a_{mi}(k) \neq 0, \dots, a_{jm}(k) \neq 0$ or $a_{mj}(k) \neq 0, \dots, a_{im}(k) \neq 0$	$\Gamma_{ij}(f) \neq 0$
Spurious coupling		$y_m \Rightarrow y_i$ and $y_m \Rightarrow y_j$	$a_{sm}(k) \neq 0, \dots, a_{is}(k) \neq 0$ and $a_{sm}(k) \neq 0, \dots, a_{js}(k) \neq 0$	

to the study of coupling, and from bivariate ($M = 2$) to MV ($M \geq 3$) processes. Specifically, our definition of direct causality agrees with the Granger's original statement [39] for bivariate processes, and with the notion of *prima facie* Granger causality introduced later in [40] for multivariate processes. The definition of causality is a generalization incorporating both direct and indirect causal influences from one process to another, while the coupling definitions generalize the causality definitions by accounting for both forward and backward interactions.

In addition to the definitions provided above, we state the following definitions of coupling, which are referred to as spurious because they concern a mathematical formalism rather than an intuitive property of two interacting processes: *spurious direct coupling* between y_i and y_j exists if $y_i \rightarrow y_m$ and $y_j \rightarrow y_m$ for at least one $m \in \{1, \dots, M\}$, $m \neq i, m \neq j$; *spurious coupling* between y_i and y_j exists if $y_m \Rightarrow y_i$ and $y_m \Rightarrow y_j$ for at least one $m \in \{1, \dots, M\}$, $m \neq i, m \neq j$. These definitions suggest that two processes can be interpreted as directly coupled also when they both directly cause a third common process, and as coupled also when they are both caused by a third common process, respectively, and are introduced here to provide a formalism for explaining a confounding property of the two common frequency-domain coupling measures reviewed in Section 3.

An illustrative example of the described causality and coupling relations is reported in Figure 1, showing a network of $M = 5$ interacting processes where each node represents a process and the connecting arrows represent coupling or causality relations. The structure of the process is unambiguously determined by the direct causality relations set in Figure 1(a), that is, $y_1 \rightarrow y_2$, $y_2 \rightarrow y_3$, $y_3 \rightarrow y_4$, $y_4 \rightarrow y_2$, and $y_1 \rightarrow y_5$. All other connectivity definitions can be established from this set of direct causality effects. Indeed, the causality relations follow from the presence of either direct causality (Figure 1(b), black arrows) or indirect causality (Figure 1(b), red arrows). Direct coupling exists as a consequence of direct causality (Figure 1(c), solid arrows), and also as a consequence of the common driving exerted by y_1 and y_4 on y_2 , such that the spurious connection $y_1 \leftrightarrow y_4$ (Figure 1(c), dashed arrow) arises. Finally, coupling is

detected between each pair of processes: while most relations derive from the causality effects (Figure 1(d), solid arrows), the relations $y_2 \Leftrightarrow y_5$, $y_3 \Leftrightarrow y_5$, and $y_4 \Leftrightarrow y_5$ are spurious as they derive from the common driving exerted by y_1 on y_2 and y_5 , on y_3 and y_5 , and on y_4 and y_5 (Figure 1(d), dashed arrows).

2.2. Multivariate Autoregressive Processes. In the linear signal processing framework, the MV closed-loop process $\mathbf{Y}(n) = [y_1(n), \dots, y_M(n)]^T$ can be represented as the output of a MV linear shift-invariant filter [15]:

$$\mathbf{Y}(n) = \sum_{k=-\infty}^{\infty} \mathbf{H}(k) \mathbf{U}(n-k), \quad (2)$$

where $\mathbf{U}(n) = [u_1(n) \dots u_M(n)]^T$ is a vector of M zero-mean input processes and $\mathbf{H}(k)$ is the $M \times M$ filter impulse response matrix. A very common representation of (2), extensively used in time series analysis, is the MV autoregressive (MVAR) representation [15]:

$$\mathbf{Y}(n) = \sum_{k=1}^p \mathbf{A}(k) \mathbf{Y}(n-k) + \mathbf{U}(n), \quad (3)$$

where $\mathbf{A}(k)$ are $M \times M$ coefficient matrices in which the element $a_{ij}(k)$ describes the dependence of $y_i(n)$ on $y_j(n-k)$ ($i, j = 1, \dots, M; k = 1, \dots, p$). Note that (3) is a particularization of (1) in which each function f_m is a linear first-order polynomial. The input process $\mathbf{U}(n)$, also called innovation process, is assumed to be composed of white and uncorrelated noises; this means that the correlation matrix of $\mathbf{U}(n)$, $\mathbf{R}_U(k) = E[\mathbf{U}(n)\mathbf{U}^T(n-k)]$, is zero for each lag $k > 0$, while it is equal to the covariance matrix $\Sigma = \text{cov}(\mathbf{U}(n))$ for $k = 0$. Under the assumption of strict causality, the input white noises are uncorrelated even at lag zero, so that their covariance reduces to the diagonal matrix $\Sigma = \text{diag}(\sigma^2_i)$.

One major benefit of the representation in (3) is that it allows to investigate properties of the joint description of the processes y_m from the model coefficients. In fact, the connectivity definitions provided in Section 2.1 for a general

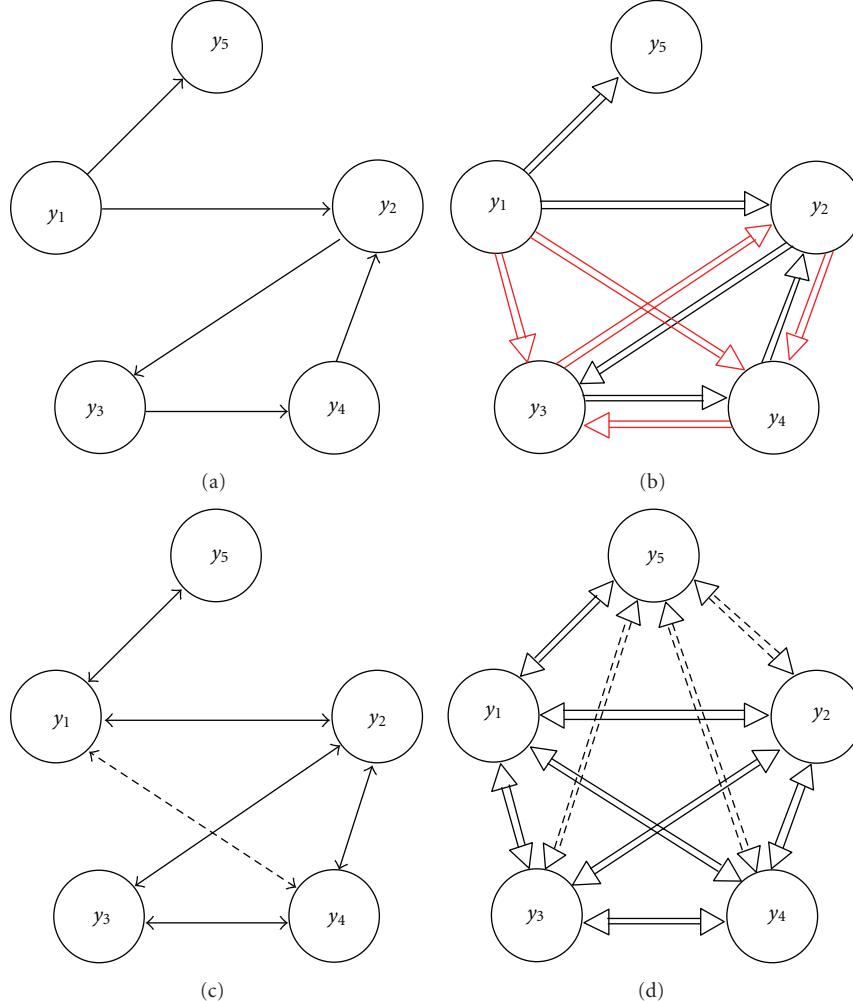


FIGURE 1: Graphical models for an illustrative five-dimensional closed loop process, denoting the scalar processes ($y_i, i = 1, \dots, 5$) as graph nodes and the connectivity relations between processes as connecting arrows. Graphs depict an imposed set of direct causality relations ($y_j \rightarrow y_i$, (a)), as well as the corresponding sets of causality ($y_j \Rightarrow y_i$, (b)), direct coupling ($y_i \leftrightarrow y_j$, (c)), and coupling ($y_i \Leftrightarrow y_j$, (d)) relations. Indirect causality relations are depicted with red arrows in (b), while spurious direct coupling and spurious coupling relations are depicted with dashed double-head arrows in ((c) and (d)).

closed-loop MV process can be specified for an MVAR process in terms of the elements of $\mathbf{A}(k)$. Conceptually, causality and coupling relations are found when the pathway relevant to the interaction is active, that is, is described by nonzero coefficients in \mathbf{A} (see Table 1). More formally, we have that $y_j \rightarrow y_i$ if $a_{ij}(k) \neq 0$ for at least one $k \in \{1, \dots, p\}$; $y_j \Rightarrow y_i$ if $a_{m_l m_{l-1}}(k_l) \neq 0$ for at least one set of $L + 1$ different values for $m_l \in \{1, \dots, M\}$ with $m_0 = j$, $m_L = i$, and one set of L lags $k_l \in \{1, \dots, p\}$ ($l = 1, \dots, L; 1 \leq L < M$); $y_i \leftrightarrow y_j$ if, for at least one pair $k_1, k_2 \in \{1, \dots, p\}$, one of the following holds: (i) $a_{ji}(k_1) \neq 0$ or $a_{ij}(k_2) \neq 0$ (direct coupling), or (ii) $a_{mi}(k_1) \neq 0$ and $a_{mj}(k_2) \neq 0$ for at least one $m \in \{1, \dots, M\}$ such that $m \neq i, m \neq j$ (spurious direct coupling); $y_i \Leftrightarrow y_j$ if, for some $m_l \in \{1, \dots, M\}$ and $k_l \in \{1, \dots, p\}$ one of the following holds: (i) $a_{m_l m_{l-1}}(k_l) \neq 0$ with either $m_0 = i, m_L = j$ or $m_0 = j, m_L = i$ (coupling), or (ii) $a_{m_l m_{l-1}}(k_l) \neq 0$ with both $m_0 = m, m_L = i$ and $m_0 = m, m_L = j$.

$m_L = j$ for at least one $m \in \{1, \dots, M\}$ such that $m \neq i, m \neq j$ (spurious coupling).

To illustrate these time-domain connectivity definitions, let us consider the MVAR process of dimension $M = 5$ and order $p = 2$:

$$\begin{aligned}
 y_1(n) &= 2\rho_1 \cos(2\pi f_1) y_1(n-1) - \rho_1^2 y_1(n-2) + u_1(n), \\
 y_2(n) &= 0.5 y_1(n-1) + 0.5 y_4(n-1) + u_2(n), \\
 y_3(n) &= 0.5 y_2(n-1) + 0.5 y_2(n-2) + u_3(n), \\
 y_4(n) &= 2\rho_4 \cos(2\pi f_4) y_4(n-1) - \rho_4^2 y_4(n-2) \\
 &\quad + 0.5 y_3(n-1) + 0.5 y_3(n-2) + u_4(n), \\
 y_5(n) &= 0.5 y_1(n-1) + 0.5 y_1(n-2) + u_5(n),
 \end{aligned} \tag{4}$$

with $\rho_1 = 0.9$, $f_1 = 0.1$, $\rho_4 = 0.8$, $f_4 = 0.3$, where the inputs $u_i(n)$ are fully uncorrelated and with variance

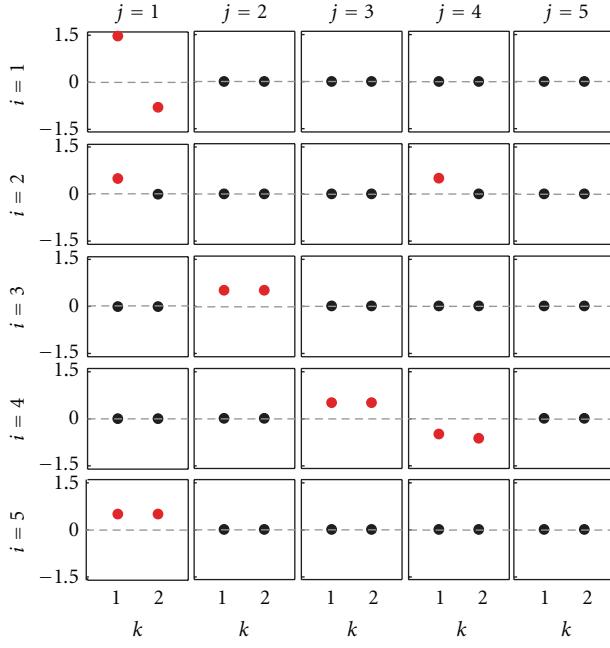


FIGURE 2: Time-domain connectivity pattern for the illustrative MVAR process of (4). Each plot depicts the values set for the coefficients $a_{ij}(k)$ ($i, j = 1, \dots, 5; k = 1, 2$), with nonzero coefficients evidenced in red.

$\sigma^2_i = 1$ ($i = 1, \dots, 5$). Equation (4) defines one of the possible MVAR processes realizing the connectivity patterns depicted in Figure 1. The matrix layout plot of Figure 2, depicting the values set for the coefficients $a_{ij}(k)$, provides a straightforward interpretation of connectivity in the time domain. In fact, non-zero values in the coefficient matrices $\mathbf{A}(1)$ and $\mathbf{A}(2)$ determine direct causality and causality among the processes—and consequently direct coupling and coupling—in agreement with the definitions provided above. In particular, we note that direct causality from y_j to y_i occurs if at least one coefficient in the (i, j) th plot of the matrix layout of Figure 2 is nonzero (red symbols). For example, nonzero values of $a_{21}(1)$ and of $\{a_{32}(1), a_{32}(2)\}$ determine the direct causality relations $y_1 \rightarrow y_2$ and $y_2 \rightarrow y_3$ when considered separately, as well as the causality relation $y_2 \Rightarrow y_3$ (indirect effect) when considered together; nonzero values of $a_{21}(1)$ and of $a_{24}(1)$ determine the direct coupling relations $y_1 \leftrightarrow y_2$ and $y_2 \leftrightarrow y_4$, and also the spurious direct coupling $y_1 \leftrightarrow y_4$; nonzero values of $\{a_{21}(1), a_{32}(1), a_{32}(2)\}$ and of $\{a_{51}(1), a_{51}(2)\}$ determine the coupling relations $y_1 \leftrightarrow y_3$ and $y_1 \leftrightarrow y_5$, but also the spurious coupling $y_3 \leftrightarrow y_5$. Note that the diagonal values of $\mathbf{A}(k)$ do not provide direct information on connectivity, but rather determine autonomous oscillations in the processes. In this case, narrow-band oscillations are generated for the process y_i by setting complex-conjugate poles with modulus ρ_i and phases $\pm 2\pi f_i$ (i.e., imposing $a_{ii}(1) = 2\rho_i \cos(2\pi f_i)$ and $a_{ii}(2) = -\rho_i^2$, $i = 1, 4$).

3. Connectivity Definitions in the Frequency Domain

3.1. Connectivity Measures. The derivation of connectivity measures which reflect and quantify in the frequency domain the time-domain definitions provided in Section 2 proceeds in two steps: first, the known correlation and partial correlation time-domain analyses are transposed in the frequency domain to describe the concepts of coupling and direct coupling, respectively; second, the parametric representation of the process is exploited to decompose the derived spectral measures of (direct) coupling into measures of (direct) causality. As to the first step, time-domain interactions within the MV closed-loop process $\mathbf{Y}(n)$ may be characterized by means of the time-lagged correlation matrix $\mathbf{R}(k) = E[\mathbf{Y}(n)\mathbf{Y}^T(n-k)]$ and of its inverse $\mathbf{R}(k)^{-1}$, whose elements may be used to define the so called correlation coefficient and partial correlation coefficient between two processes y_i and y_j [41]:

$$\rho_{ij}(k) = \frac{r_{ij}(k)}{\sqrt{r_{ii}(k)r_{jj}(k)}}, \quad (5)$$

$$\eta_{ij}(k) = -\frac{p_{ij}(k)}{\sqrt{p_{ii}(k)p_{jj}(k)}},$$

where $r_{ij}(k)$ and $p_{ij}(k)$ are the $i-j$ elements of $\mathbf{R}(k)$ and $\mathbf{R}(k)^{-1}$. The correlation ρ_{ij} is a normalized measure of the linear interdependence between $y_i(n)$ and $y_j(n-k)$ and, as such, quantifies coupling in the time-domain. The partial correlation η_{ij} is a measure of direct coupling, in the sense that it quantifies the linear interdependence between $y_i(n)$ and $y_j(n-k)$ after removing the effects of all remaining processes, according to a procedure denoted as partialization [42]. The frequency-domain counterpart of these measures is obtained considering the traditional spectral analysis of MV processes on one side [15], and the corresponding dual analysis performed in the inverse spectral domain on the other side [43]. Specifically, the $M \times M$ spectral density matrix $\mathbf{S}(f)$ is defined as the Fourier Transform (FT) of $\mathbf{R}(k)$, while the inverse spectral matrix $\mathbf{P}(f) = \mathbf{S}(f)^{-1}$ results as the FT of the partial correlation matrix $\mathbf{R}(k)^{-1}$. The elements of the spectral matrices $\mathbf{S}(f)$ and $\mathbf{P}(f)$ are combined to define the so-called *coherence* (Coh) and *partial coherence* (PCoh) functions [44]:

$$\Gamma_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}}, \quad (6a)$$

$$\Pi_{ij}(f) = -\frac{P_{ij}(f)}{\sqrt{P_{ii}(f)P_{jj}(f)}}. \quad (6b)$$

When a closed-loop MV process is particularized to a MVAR process, the spectral representation may be obtained taking the FT of the representations in (2) and (3), which yields $\mathbf{Y}(f) = \mathbf{H}(f)\mathbf{U}(f)$ and $\mathbf{Y}(f) = \mathbf{A}(f)\mathbf{Y}(f) + \mathbf{U}(f)$, respectively, where $\mathbf{Y}(f)$ and $\mathbf{U}(f)$ are the FTs of $\mathbf{Y}(n)$ and

$\mathbf{U}(n)$, and the $M \times M$ transfer matrix and coefficient matrix are defined in the frequency domain as:

$$\mathbf{H}(f) = \sum_{k=-\infty}^{\infty} \mathbf{H}(k) e^{-j2\pi fkT}, \quad \mathbf{A}(f) = \sum_{k=1}^p \mathbf{A}(k) e^{-j2\pi fkT}. \quad (7)$$

Comparing the two spectral representations, it is easy to show that the coefficient and transfer matrices are linked by: $\mathbf{H}(f) = [\mathbf{I} - \mathbf{A}(f)]^{-1} = \overline{\mathbf{A}}(f)^{-1}$. This important relation is useful to draw the connection between the spectral density matrix $\mathbf{S}(f)$ and its inverse $\mathbf{P}(f)$, as well as to decompose the symmetric frequency-domain connectivity measures into terms eliciting directionality. The key element is the spectral factorization theorem [45]:

$$\mathbf{S}(f) = \mathbf{H}(f) \Sigma \mathbf{H}^H(f), \quad \mathbf{P}(f) = \overline{\mathbf{A}}^H(f) \Sigma^{-1} \overline{\mathbf{A}}(f), \quad (8)$$

(where H stands for Hermitian transpose), which allows to represent, under the assumption of strict causality, the elements of the spectral density matrices as:

$$\begin{aligned} S_{ij}(f) &= \sum_{m=1}^M \sigma_m^2 H_{im}(f) H_{jm}^*(f), \\ P_{ij}(f) &= \sum_{m=1}^M \frac{1}{\sigma_m^2} \overline{A}_{mi}^*(f) \overline{A}_{mj}(f). \end{aligned} \quad (9)$$

The spectral decompositions in (9) lead to decompose the Coh and PCoh defined in (6a) and (6b) as:

$$\Gamma_{ij}(f) = \sum_{m=1}^M \frac{\sigma_m H_{im}(f)}{\sqrt{S_{ii}(f)}} \frac{\sigma_m H_{jm}^*(f)}{\sqrt{S_{jj}(f)}} = \sum_{m=1}^M \gamma_{im}(f) \gamma_{jm}^*(f), \quad (10a)$$

$$\begin{aligned} \Pi_{ij}(f) &= - \sum_{m=1}^M \frac{(1/\sigma_m) \overline{A}_{mj}(f)}{\sqrt{P_{jj}(f)}} \frac{(1/\sigma_m) \overline{A}_{mi}^*(f)}{\sqrt{P_{ii}(f)}} \\ &= - \sum_{m=1}^M \pi_{mj}(f) \pi_{mi}^*(f). \end{aligned} \quad (10b)$$

The last terms in (10a) and (10b) contain, respectively, the so-called *directed coherence* (DC) and *partial directed coherence* (PDC), which we define in this study as:

$$\gamma_{ij}(f) = \frac{\sigma_j H_{ij}(f)}{\sqrt{\sum_{m=1}^M \sigma_m^2 |H_{im}(f)|^2}}, \quad (11a)$$

$$\pi_{ij}(f) = \frac{(1/\sigma_i) \overline{A}_{ij}(f)}{\sqrt{\sum_{m=1}^M (1/\sigma_m^2) |\overline{A}_{mj}(f)|^2}}. \quad (11b)$$

The DC as defined in (11a) was originally proposed by Saito and Harashima [46], and further developed as connectivity measure by Baccala et al. [18]. Note that the directed transfer function (DTF) defined in [12] is a particularization of the DC in which the input variances are all equal ($\sigma^2_1 = \dots = \sigma^2_M$) so that they cancel each other in (11a). The quantity which we define as PDC in (11b) was named “generalized PDC” in [20], while the original version of the PDC [19] was not including inner normalization by the input noise variances; our definition (11b) follows directly from the decomposition in (10b). We note that other variants of the PDC estimator have been recently provided: the “information PDC” [21], which under the hypothesis of strict causality reduces to (11a), has been proposed as a measure bridging frequency and information domains; the “renormalized PDC” [22] has been proposed to allow drawing conclusion about the interaction strength by normalization. Here, besides the meaningful dual derivation of DC and PDC as factors in the decomposition of Coh and PCoh, we further justify the utilization of the measures defined in (11a) and (11b) noting that they satisfy the desirable property of scale-invariance. On the contrary, as shown by Winterhalder et al. [11], false detections of causality may occur from low variance process to processes with significantly higher variance when the original DTF and PDC estimators are used.

The quantities $\gamma_{ij}(f)$ and $\pi_{ij}(f)$ defined in (11a) and (11b) can be interpreted as measures of the influence of y_j onto y_i , as opposed to $\gamma_{ji}(f)$ and $\pi_{ji}(f)$ which measure the interaction occurring over the opposite direction from y_i onto y_j . Therefore, the DC and the PDC, being factors in the decomposition of Coh and PCoh, are asymmetric connectivity measures which elicit the directional information from the two symmetric measures. More detailed interpretation of all these measures is provided in the next subsection.

3.2. Interpretation. A straightforward interpretation of the four connectivity measures above presented may be obtained considering that they reflect in the frequency domain the different time-domain definitions of connectivity given in Section 2.2. First, we note that the PDC is a measure of direct causality, because the numerator of (11b) contains, with $i \neq j$, the term $\overline{A}_{ij}(f)$, which is nonzero only when $a_{ij}(k) \neq 0$ for some k and is uniformly zero when $a_{ij}(k) = 0$ for each k . Considering the DC, one can show that, expanding $\mathbf{H}(f) = \overline{\mathbf{A}}(f)^{-1}$ as a geometric series [36], the transfer function $H_{ij}(f)$ contains a sum of terms each one related to one of the (direct or indirect) transfer paths connecting y_j to y_i ; therefore, the numerator of (11a) is nonzero whenever any path connecting y_j to y_i is significant, that is, when causality occurs from y_j to y_i . As to the coupling definitions, we note from (10a) and (10b) that $\Gamma_{ij}(f) \neq 0$ when both $\gamma_{im}(f) \neq 0$ and $\gamma_{jm}(f) \neq 0$, and $\Pi_{ij}(f) \neq 0$ when both $\pi_{mi}(f) \neq 0$ and $\pi_{mj}(f) \neq 0$; this suggests that Coh and PCoh reflect respectively coupling and direct coupling relations in accordance with a frequency-domain representation of the definitions given in Section 2. However, (10a) and (10b) explain also the rationale of introducing a mathematical formalism to define spurious direct coupling and spurious coupling. In fact, the fulfillment of $\Pi_{ij}(f) \neq 0$ or $\Gamma_{ij}(f) \neq 0$ at a given frequency f is not a sufficient condition for the existence of direct coupling or coupling at that frequency, because the observed relation can be also spurious. The correspondence with the

time-domain connectivity definitions and the frequency-domain measures is summarized in Table 1.

The four connectivity measures defined in (6a), (6b), (11a), and (11b) are complex-valued. In order to have real-valued measures, the squared modulus of Coh, PCoh, DC, and PDC is commonly used to measure connectivity in the frequency domain. Therefore, $|\Gamma_{ij}(f)|^2$, $|\Pi_{ij}(f)|^2$, $|\gamma_{ij}(f)|^2$, and $|\pi_{ij}(f)|^2$ are computed to quantify, respectively, coupling, direct coupling, causality, and direct causality as a function of frequency. All these squared measures are normalized so that they take values between 0, representing absence of connectivity, and 1, representing full connectivity between the processes y_i and y_j at the frequency f . This property allows to interpret the value of each squared index as a measure of the “strength” of connectivity. While this interpretation is meaningful for Coh and DC, it is less useful for PCoh and PDC. Indeed, Coh and DC are defined from the elements of the spectral matrix $\mathbf{S}(f)$ (see (6a)–(11a)) and, as such, are easy to interpret in terms of power spectral density. On the contrary, PCoh and PDC are obtained from a dual representation evidencing the inverse spectral matrix $\mathbf{P}(f)$ (see (6b)–(11b)), which is less easy to interpret because inverse spectra do not have a clear physical meaning. On the other hand, PCoh and PDC are useful when one is interested in determining the frequency-domain connectivity structure of a vector process, as they elicit direct connections between two processes in the MV representation. From this point of view, Coh and DC are more confusing as they measure the “total” connectivity between two processes, mixing together direct and indirect effects.

Further interpretation of the directional measures of connectivity, is provided considering the normalization properties $\sum_{m=1}^M |\gamma_{im}(f)|^2 = 1$ and $\sum_{m=1}^M |\pi_{mj}(f)|^2 = 1$, indicating that the DC is normalized with respect to the structure that receives the signal and the PDC is normalized with respect to the structure that sends the signal [19]. Combined with (9), these properties lead to represent the spectra and inverse spectra of a scalar process, that is, the diagonal elements of $\mathbf{S}(f)$ and $\mathbf{P}(f)$, as:

$$S_{ii}(f) = \sum_{m=1}^M S_{i|m}(f), \quad S_{i|m}(f) = |\gamma_{im}(f)|^2 S_{ii}(f), \quad (12a)$$

$$P_{jj}(f) = \sum_{m=1}^M P_{j-m}(f), \quad P_{j-m}(f) = |\pi_{mj}(f)|^2 P_{jj}(f), \quad (12b)$$

where $S_{i|m}(f)$ is the part of $S_{ii}(f)$ due to y_m , and $P_{j-m}(f)$ is the part of $P_{jj}(f)$ directed to y_m . Thus, the DC and the PDC may be viewed as the relative amount of power of the output process which is received from the input process, and the relative amount of inverse power of the input process which is sent to the output process, respectively. Again, inverse power quantifies direct causality but is of difficult physical interpretation, while power is straightforward to interpret but includes both direct and indirect effects. Therefore, a

desirable development would be to split the DC into direct and indirect contributions, in order to exploit the advantages of both representations. However, such a development is not trivial, as recently shown by Gigi and Tangirala [32] who elicited the presence of an interference term which prevents the separation of direct and indirect energy transfer between two variables of a MV process.

To compare the behavior of the presented connectivity measures and to discuss their properties, let us consider the frequency-domain representation of the theoretical example with time-domain representation given by (4). The trends of spectral and cross-spectral density functions are depicted in Figure 3. The spectra of the five processes, reported as diagonal plots in Figure 3(a), exhibit clear peaks at the frequency of the two oscillations imposed at ~ 0.1 Hz and ~ 0.3 Hz for y_1 and y_4 , respectively, and appear also in the spectra of the remaining processes according to the imposed causal information transfer. On the contrary the inverse spectra, reported as diagonal plots in Figure 3(b), do not provide clear information about such an oscillatory activity. Off-diagonal plots of Figures 3(a) and 3(b) depict, respectively, the squared magnitudes of Coh and PCoh; note the symmetry of the two functions $(\Gamma_{ij}(f) = \Gamma_{ji}^*(f), \Pi_{ij}(f) = \Pi_{ji}^*(f))$, reflecting the fact that they cannot account for directionality of the considered interaction. The comparison of Figure 3(a) with Figure 1(d), and of Figure 3(b) with Figure 1(c), evidences how Coh and PCoh provide a spectral representation of coupling and direct coupling: indeed, the frequency-domain functions are uniformly zero when the corresponding connectivity relation is absent in the time domain. Note that the spurious direct coupling and spurious coupling connections evidenced in Figures 1(c) and 1(d) cannot be pointed out from the frequency-domain representations of Figure 3: for example, the coherence between y_5 and all other processes is very high at ~ 0.1 Hz although only the coupling $y_1 \leftrightarrow y_5$ is not spurious. Another observation regards interpretability of the absolute values: while Coh shows clear peaks at the frequency of coupled oscillations (~ 0.1 Hz and ~ 0.3 Hz) when relevant, PCoh is less easy to interpret as sometimes the squared modulus does not exhibit clear peaks (e.g., $|\Pi_{23}(f)|^2, |\Pi_{34}(f)|^2$) or is very low (e.g., $|\Pi_{12}(f)|^2, |\Pi_{14}(f)|^2$).

Figure 4 depicts the spectral decomposition of the MVAR process, as well as the trends resulting for the DC function from this decomposition. Note that the DC reflects the pattern of causality depicted in Figure 1(b), being uniformly zero along all directions over which no causality is imposed in the time domain. Figure 4(a) provides a graphical representation of (12a), showing how the spectrum of each process can be decomposed into power contributions related to all processes; normalizing these contributions, one gets the squared modulus of the DC, as depicted in Figure 4(b). In the example, the spectrum of y_1 is decomposed in one part only, deriving from the same process. This indicates that no part of the power of y_1 is due to the other processes. The absence of external contributions is reflected by the null profiles of the squared DCs $|\gamma_{1j}(f)|^2$ for each $j > 1$, which also entail a flat unitary profile for $|\gamma_{11}(f)|^2$. On the

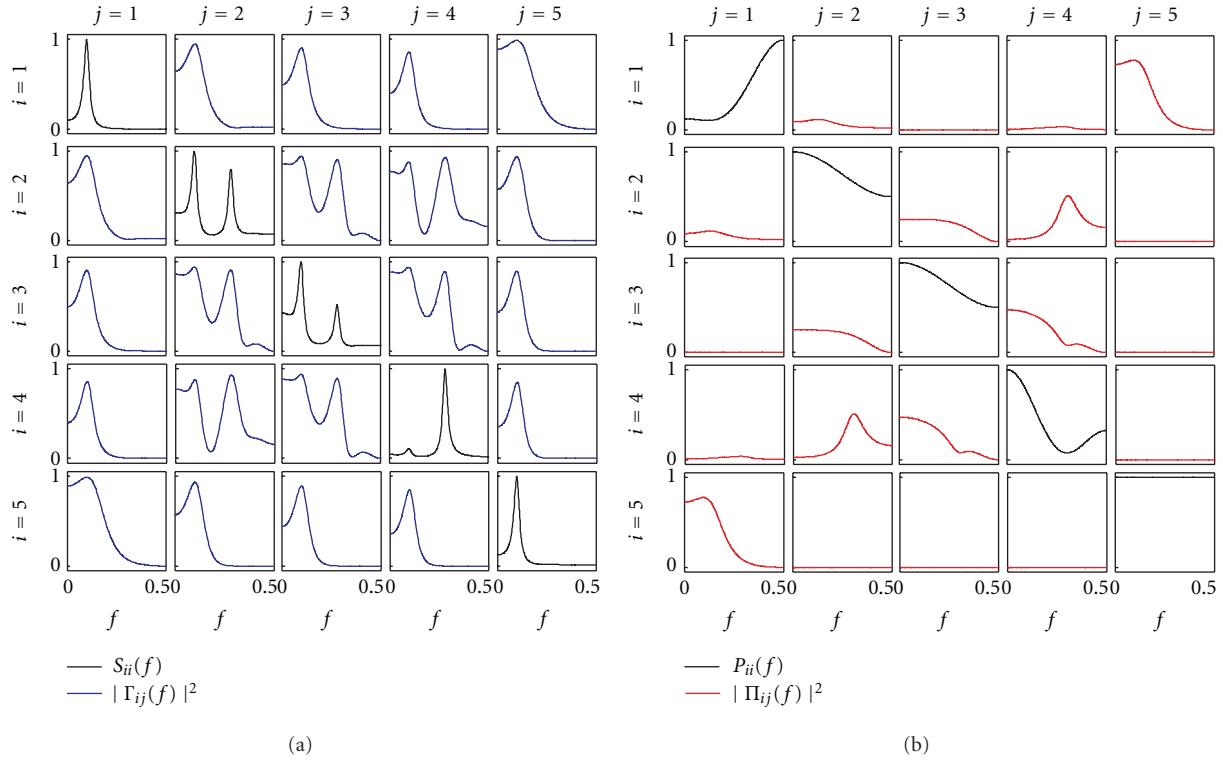


FIGURE 3: Spectral functions and frequency domain measures of coupling for the illustrative MVAR process of (4). (a) Power spectral density of the process y_i ($S_{ii}(f)$, black) and coherence between y_i and y_j ($|\Gamma_{ij}(f)|^2$, blue). (b) Inverse power spectral density of y_i ($P_{ii}(f)$, black) and partial coherence between y_i and y_j ($|\Pi_{ij}(f)|^2$, red).

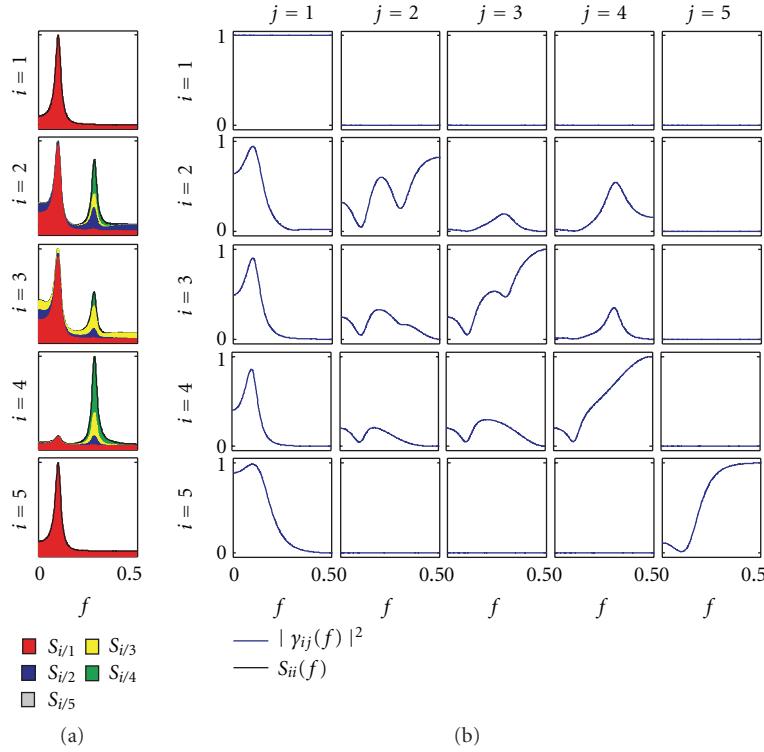


FIGURE 4: Decomposition of the power spectrum of each process y_i in (4), $S_{ii}(f)$, into contributions coming from each process y_j (S_{ij} , shaded areas in each plot) (a), and corresponding squared DC from y_j to y_i , $|\gamma_{ij}(f)|^2$ (b) depicted for each $i, j = 1, \dots, M$.

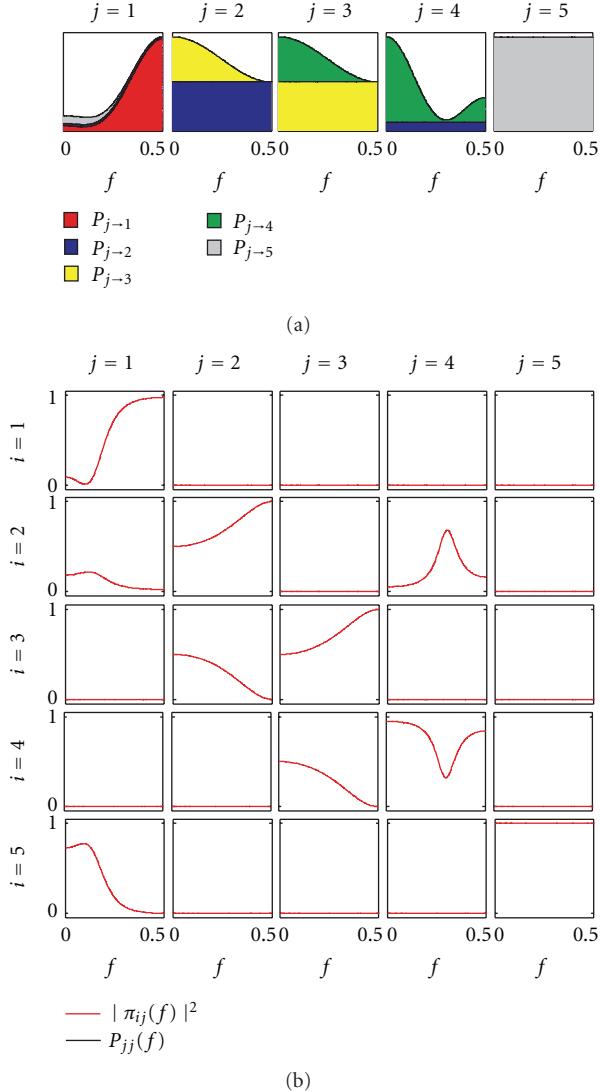


FIGURE 5: Decomposition of the inverse power spectrum of each process y_i in (4), $P_{jj}(f)$, into contributions directed towards each process y_j (P_{j-i} , shaded areas in each plot) (a), and corresponding squared PDC from y_j to y_i , $|\pi_{ij}(f)|^2$ (b) depicted for each $i, j = 1, \dots, M$.

contrary, the decompositions of y_i , with $i > 1$, results in contributions from the other processes, so that the squared DC $|\gamma_{ij}(f)|^2$ is nonzero for some $j \neq i$, and the squared DC $|\gamma_{ii}(f)|^2$ is not uniformly equal to 1 as a result of the normalization condition. In particular, we note that the power of the peak at $f_1 = 0.1$ Hz is due to y_1 for all processes (red areas in Figure 4(a)), determining very high values of the squared DC in the first column of the matrix plot in Figure 3(b), that is, $|\gamma_{11}(f_1)|^2 \approx 1$; this behavior represents in the frequency domain the causality relations imposed from y_1 to all other processes. Note that, as a consequence of the normalization condition of the DC, the high values measured at ~ 0.1 Hz for $|\gamma_{11}|^2$ entail very low values, at the same frequency, for $|\gamma_{ij}|^2$ computed with $j > 1$.

Whereas this property is straightforward when the studied effect is direct, in the case of indirect causality, it suggests that the DC modulus tends to ascribe the measured causal coupling to the source process (i.e., the first process) rather than to the intermediate processes of the considered cascade interaction. The remaining causality relations are relevant to the oscillation at $f_2 = 0.3$ Hz, which is generated in y_4 and then re-transmitted to the same process through a loop involving y_2 and y_3 . This loop of directed interactions is reflected by the presence of a peak at ~ 0.3 Hz of the spectra of y_2 , y_3 , and y_4 , as well as by the spectral decomposition within this frequency band (Figure 4(a)). This decomposition results, after proper normalization, in the nonzero DCs $|\gamma_{42}|^2$, $|\gamma_{23}|^2$, $|\gamma_{34}|^2$ (direct causality) and $|\gamma_{43}|^2$, $|\gamma_{32}|^2$, $|\gamma_{24}|^2$ (indirect causality) observed at f_2 .

A dual interpretation to that provided above holds for the decomposition of the inverse spectra into absolute and normalized contributions sent to all processes, which are depicted for the considered example in the area plot of Figure 5(a) and in the matrix PDC plot of Figure 5(b), respectively. The difference is that now contributions are measured as outflows instead as inflows, are normalized to the structure sending the signal instead to that receiving the signal, and reflect the concept of direct causality instead that of causality. With reference to the proposed example, we see that the inverse spectrum of y_1 is decomposed into contributions flowing out towards y_2 and y_5 (blue and gray areas underlying $P_{11}(f)$ in Figure 5(a)), which are expressed in normalized units by the squared PDCs $|\pi_{21}|^2$ and $|\pi_{51}|^2$. While y_2 , y_3 , and y_5 interact in a closed loop (absolute units: $P_{2-3} \neq 0$, $P_{3-4} \neq 0$, $P_{4-2} \neq 0$; normalized units: $|\pi_{32}|^2 \neq 0$, $|\pi_{43}|^2 \neq 0$, $|\pi_{24}|^2 \neq 0$), y_5 does not send information to any process ($P_{5-i} = 0$, $|\pi_{i5}|^2 = 0$, $i = 1, 2, 3, 4$). As can be seen comparing Figure 5 with Figure 1(a), the profiles of P_{j-i} and $|\pi_{ij}|^2$ provide a frequency-domain description, respectively, in absolute and normalized terms, of the imposed pattern of direct causality. We note also that all inverse spectra of a process contain a contribution coming from the same process, which describes the part of $P_{jj}(f)$ which is not sent to any of the other processes (P_{j-j} in Figure 4(a)). After normalization, this contribution is quantified by the PDC $|\pi_{jj}|^2$, as depicted by the diagonal plots of Figure 5(b).

4. Practical Analysis

4.1. Model Estimation. The practical application of the theoretical concepts described in this tutorial paper is based on considering the available set of time series measured from the system under analysis, $\{y_m(n), m = 1, \dots, M; n = 1, \dots, N\}$, as a finite-length realization of the MV stochastic process describing the evolution of the system over time. Hence, the descriptive equation (3) is seen as a model of how the observed data have been generated, and model identification algorithms have to be applied for providing estimates of the model coefficients to be used in place of the true unknown coefficients in the generating equations. Obviously, the estimates will never be the exact coefficients,

and consequently the frequency-domain measures estimated from the real data will always be an approximation of the true functions. The goodness of the approximation depends on practical factors such as the data length, and on the type and parameters of the procedure adopted for the identification of the model coefficients. Identification of the MVAR model (3) can be performed with relative ease by means of estimation methods based on the principle of minimizing the prediction error, that is, the difference between actual and predicted data (see, e.g., [15] or [47] for detailed descriptions). A simple estimator is the MV least-squares method, which is based first on representing (3) in compact representation as $\mathbf{Y} = \mathbf{AZ} + \mathbf{U}$, where $\mathbf{A} = [\mathbf{A}(1) \cdots \mathbf{A}(p)]$ is the $M \times pM$ matrix of the unknown coefficients, $\mathbf{Y} = [\mathbf{Y}(p+1) \cdots \mathbf{Y}(N)]$ and $\mathbf{U} = [\mathbf{U}(p+1) \cdots \mathbf{U}(N)]$ are $M \times (N-p)$ matrices, and $\mathbf{Z} = [\mathbf{Z}_1^T \cdots \mathbf{Z}_p^T]^T$ is a $pM \times (N-p)$ matrix having $\mathbf{Z}_i = [\mathbf{Y}(p-i+1) \cdots \mathbf{Y}(N-i)]$ as i th row block ($i = 1, \dots, p$). The method estimates the coefficient matrices through the well-known least-squares formula: $\hat{\mathbf{A}} = \mathbf{YZ}^T[\mathbf{ZZ}^T]^{-1}$, and the innovation process as the residual time series: $\hat{\mathbf{U}} = \hat{\mathbf{AZ}} - \mathbf{Y}$. As to model order selection, several criteria exist for its determination [47]. Common approaches are to follow the Akaike information criterion (AIC, [48]) or to the Bayesian information criterion (BIC, [49]), based on setting the order p at the value for which the respective figure of merit, (i.e., $AIC(p) = N \cdot \log(\det \Sigma) + 2M^2 p$, or $BIC(p) = N \cdot \log(\det \Sigma) + \log(N)M^2 p$) reaches a minimum within a predefined range of orders. While the model identification and order selection methods presented here have good statistical properties, more accurate approaches exist; for example, we refer the reader to [50] for a comparison of different MVAR estimators, to [51] for order selection criteria optimized for MVAR models, and to [52] for an identification approach combining MVAR coefficient estimation and order selection.

After model identification, validation steps are necessary to guarantee a correct interpretation of the obtained results. Model validation refers to the use of a range of diagnostic tools which are available for checking the adequacy of the estimated structure. In particular, identification of the MVAR model (3) should result in temporally uncorrelated and mutually independent residuals $\hat{\mathbf{U}}(n)$. These assumptions may be checked, for example, using the Ljung-Box portmanteau test for whiteness and the Kendall's τ test for independence [47]. Mutual independence of the residuals has to be checked particularly at zero lag, because the existence of correlated model innovations violates the assumption of strict causality. Although these tests are often skipped in practical analysis, we stress the importance of performing model validation, because failure to satisfy the model assumptions is a clear indication of model misspecification (see Section 5 for a more detailed description of this problem).

4.2. Statistical Significance of Connectivity Measures. Besides confirming the suitability of the estimated model, another issue of great practical importance is the assessment of the statistical significance of the derived connectivity measures. Due to practical estimation problems, nonzero values are

indeed likely to occur at some frequencies even in the case of absence of a true interaction between the two considered processes. This problem is commonly faced by means of statistical hypothesis testing procedures based on setting a threshold for significance at the upper limit of the confidence interval of the considered index, where confidence intervals are based on the sampling distribution of the index computed under the null hypothesis of absence of connectivity. Comparing at each specific frequency the estimated index with the threshold allows rejection or acceptance of the null hypothesis according to the predetermined level of significance. The sampling distribution in the absence of connectivity may be derived either theoretically or empirically: theoretical approaches are elegant and computationally more efficient, empirical ones are more general and free of analytical approximations. While the statistical analytical threshold for the Coh estimator can be found in classic time series analysis books (e.g., [53]), recent theoretical studies have provided rigorous asymptotic distributions for the PDC [54, 55] and its renormalized version [22], as well as for the DC/DTF [36]. As to the determination of empirical significance levels, the most popular approaches consist in applying permutation statistics [56] when the data matrix can be partitioned in many windows from which multiple values of the connectivity measure may be computed, and in applying surrogate data analysis [57] when only one value of the measure is computed. In the latter case, the most followed approach is the generation of the so-called FT surrogate series, which are obtained by a phase randomization procedure applied independently to each series of the considered MV data set. This approach has been proposed to assess the significance of the Coh estimator [58], and has been used also with the causality estimators [17, 59]. A recent development of the FT method is that leading to the generation of the so-called "causal FT" (CFT) surrogates [60]. CFT surrogates were devised specifically for the detection of the significance of causality and direct causality in the frequency domain, and have been shown to outperform FT surrogates as regards the empirical detection of a zero-level for the DC or the PDC [60]. However, it has to be remarked that the computational burden of this new method is significantly larger than that required for the generation of FT surrogates, and this may make very demanding, or even intractable, the assessment of significance when high-dimensional MVAR models are analyzed.

4.3. Practical Illustrative Example. In this section, we report a practical application of the presented connectivity analysis to MV neurophysiological time series. Specifically, we considered electroencephalographic (EEG) recordings collected from a subject performing a visuomotor task combining precise grip motor commands with sensory visual feedback. Briefly, the subject was asked to track the variations in size of a square target displayed on a monitor by acting on a pinch grip through his right hand thumb and forefinger. Visual feedback about his performance was provided to the subject by displaying on the monitor another square

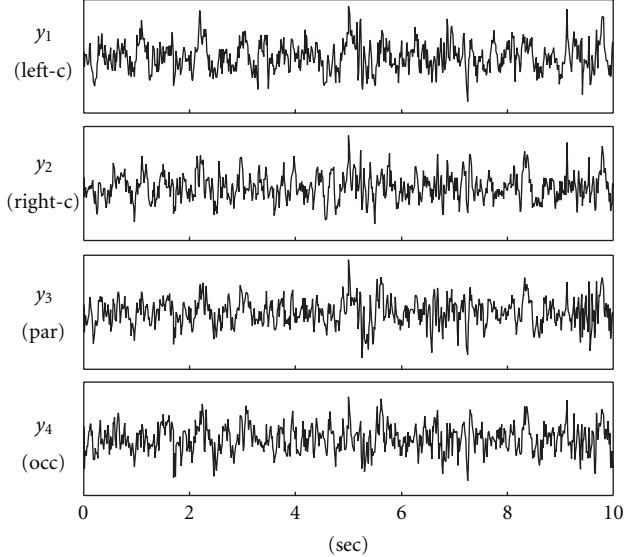


FIGURE 6: Time series considered for the neurophysiological application: EEG signals recorded at left-central (y_1), right-central (y_2), parietal (y_3), and occipital (y_4) scalp locations during the execution of a visuomotor task (see text for details).

reflecting the exerted force (the task required to continuously match the two rectangles). EEG signals were acquired (earlobes common reference; sampling rate: 576 Hz) during the experiment according to the standard 10–20 electrode placement enlarged with intermediate positions in scalp areas of interest for the specific task performed. Full details about the experimental protocol can be found in [28].

Here, we present the results of frequency-domain connectivity analysis performed for $M = 4$ EEG signals selected as representative of the cortical areas involved in visuomotor integration processes, that is, left and right central areas (motor cortex, electrodes C3 and C4) and posterior-parietal regions (visual area and parietal cortex, electrodes Pz and Oz) [61]. The signals were bandpass filtered (3–45 Hz) to remove power supply noise and extract information about the brain rhythms of interest, and then downsampled to 72 Hz to reduce redundancy. Pre-processed EEGs were carefully inspected to identify possible artifacts, and a stationary window of ten seconds ($N = 720$ samples) was then selected for the analysis. The four analyzed signals are shown in Figure 6.

The coefficients and input covariance of the MVAR model describing the four time series were estimated using the MV least-squares method; the model order, determined as the minimum of the Akaike figure of merit within the range (1,30), was $p = 8$. Model validation was performed checking whiteness and independence of the estimated model residuals by means of the Ljung-Box test and the Kendall test, respectively. The estimated coefficients and input covariance were used to estimate the frequency-domain coupling and causality functions according to (6a), (6b), (11a), and (11b), respectively. Each estimated connectivity function was evaluated inside the beta band of the frequency spectrum (13–30 Hz), in order to investigate

connectivity mechanisms related to medium-range interactions among communicating brain areas [28]. The statistical significance of the various connectivity measures computed in the beta band for each specific direction of interaction was assessed by means of an approach based on the generation of surrogate data. The test, which is described in detail in [60], was performed generating a set of 100 surrogate series by means of a phase randomization procedure that preserves the modulus of the Fourier transform of the original series and alters the Fourier phases in a way such that connectivity is destroyed only over the direction of interest; note that the method is specific for each connectivity measure, so that it specifically destroys coupling, direct coupling, causality, or direct causality between two series, respectively, when the significance of Coh, PCoh, DC, or PDC is going to be tested. For each connectivity measure, the threshold for significance was obtained as the 95th percentile (corresponding to 5% significance) of the distribution of the measure computed over the 100 surrogate series.

The results of the analysis are reported in Figures 7 and 8 for coupling and causality measures, respectively. The analysis of coupling indicates that the network of the four interacting signals is fully connected inside the beta band, as documented by the Coh values exceeding the significance threshold for each pair of time series (Figure 7(a)). When this information is particularized to the study of direct coupling through the PCoh, we observe that direct connections are set in the beta band between y_3 and y_4 , y_1 and y_3 , y_1 and y_4 , and y_1 and y_2 (though the PCoh exceeds the significance threshold only slightly in this last case). This suggests a major involvement of the left hemisphere in the connectivity network activated by the visuomotor task, likely due to the dominant role of the left-motor cortical area (signal y_1 , electrode C3); the EEG recorded from this area, which is contra-lateral to the moving right hand, is mainly linked to that recorded from the parietal (signal y_3 , electrode Pz) and occipital (signal y_4 , electrode Oz) areas which are expected to reflect processing of the visual information. The causality analysis depicted in Figure 8 shows how the information about the direction of interaction may be elicited for this application. In particular, the peaks shown inside the beta band by the squared DC computed from y_j to y_1 ($j = 2, 3, 4$; first row plots of Figure 8(a)), which are small but exceed the threshold for significance, indicate that a significant part of the power spectrum of y_1 is due to the other channels. Considering also the significant DC from y_4 to y_3 , we can infer the presence of a non negligible information transfer from the occipital to the left-central cortical regions. This finding is confirmed by the analysis of direct causality performed through the PDC (Figure 8(b)), which indicates that both the direct pathway $y_4 \rightarrow y_1$ and the indirect pathway $y_4 \rightarrow y_3 \rightarrow y_1$ are active in determining causality from the occipital to the left-central areas, as documented by the significant values of $|\pi_{14}|^2$, $|\pi_{34}|^2$, and $|\pi_{13}|^2$ inside the beta band. The unidirectional nature of the information flow is confirmed by the fact that both the DC (Figure 8(a)) and the PDC (Figure 8(b)) resulted non significant over all directions from y_j to y_i with $j < i$. Taken together, all these results suggest the existence of a functional link between

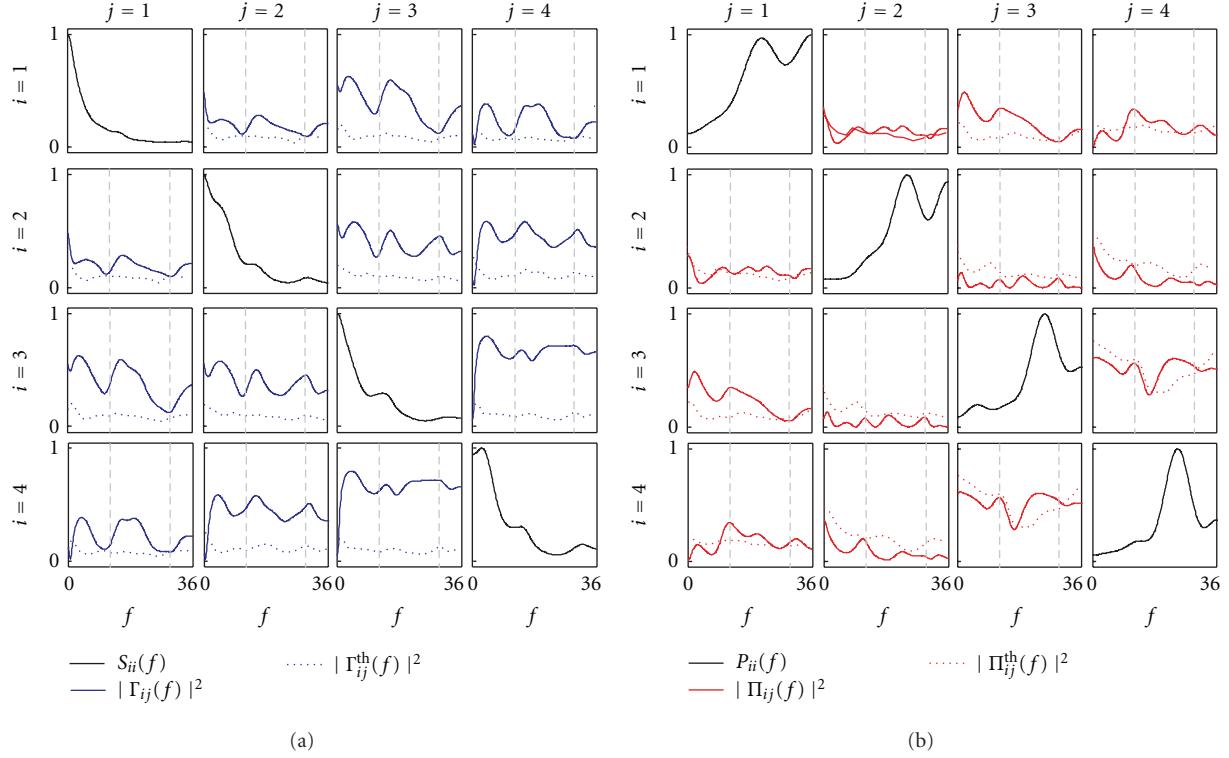


FIGURE 7: Spectral functions and frequency-domain measures of coupling for the exemplary neurophysiological application (y_1 : left central EEG; y_2 : right central EEG; y_3 : parietal EEG; y_4 : occipital EEG). (a), Power spectral density of the process $y_i(S_{ii}(f))$ and coherence between y_i and y_j ($|\Gamma_{ij}(f)|^2$) plotted together with its corresponding threshold for significance ($|\Gamma_{ij}^{th}(f)|^2$). (b), Inverse power spectral density of $y_i(P_{ii}(f))$ and partial coherence between y_i and y_j ($|\Pi_{ij}(f)|^2$) plotted together with its corresponding threshold for significance ($|\Pi_{ij}^{th}(f)|^2$). Vertical dashed lines in each off-diagonal plot denote the bounds of the beta frequency band (13–30 Hz).

motor and visual cortices during the performed visuomotor task, and lead to hypothesize an active role of the visual feedback in driving the beta oscillations measured in the motor cortex. A full analysis of this experiment, performed on more subjects and leading to a deeper interpretation of the involved sensorimotor integration mechanisms, is reported in [62].

5. Limitations and Challenges

In spite of its demonstrated usefulness and widespread utilization, MVAR-based connectivity analysis is often challenged by a number of issues that need to be taken into serious account to avoid an improper utilization of this tool. A key issue in this regard is that of model mis-specification, which occurs when the developed MVAR model does not adequately capture the correlation structure of the observed dataset. There are several factors which may determine model mis-specification, including utilization of an inappropriate model structure, incorrect model order selection, effects of non-modeled latent variables, and aspects not accountable by the traditional MVAR structure such as nonstationarity and nonlinearity. Most of these factors are typically reflected in the structure of the model residuals,

resulting in a failure for the model to fulfill the assumptions of whiteness and independence of the innovations (see Section 4.1). When the MVAR model is mis-specified, utilization of the related frequency-domain connectivity measures is potentially dangerous and is generally not recommended, because it may lead to infer misleading or inconsistent connectivity patterns, and thus to erroneously interpret the physiological mechanism under investigation. In the following, we discuss the limitations posed on MVAR-based connectivity analysis by each of the factors listed above, and we outline recent work that may address, at least in part, the related pitfalls.

5.1. Appropriateness of Model Structure. A common cause for model mis-specification is the inadequacy of the MVAR model structure to fully describe the observed set of MV time series. Validation tests (see Section 4.1) provide objective criteria on whether the model has the capability of resolving the measured dynamics and dynamical interactions. The requirements of whiteness and independence of the model residuals can be understood considering that, if the model has captured the whole temporal structure of the data, what remains after modeling (i.e., the residuals) has no temporal structure. Failure of fulfilling the white noise assumption means that the spectral properties of the signals are not

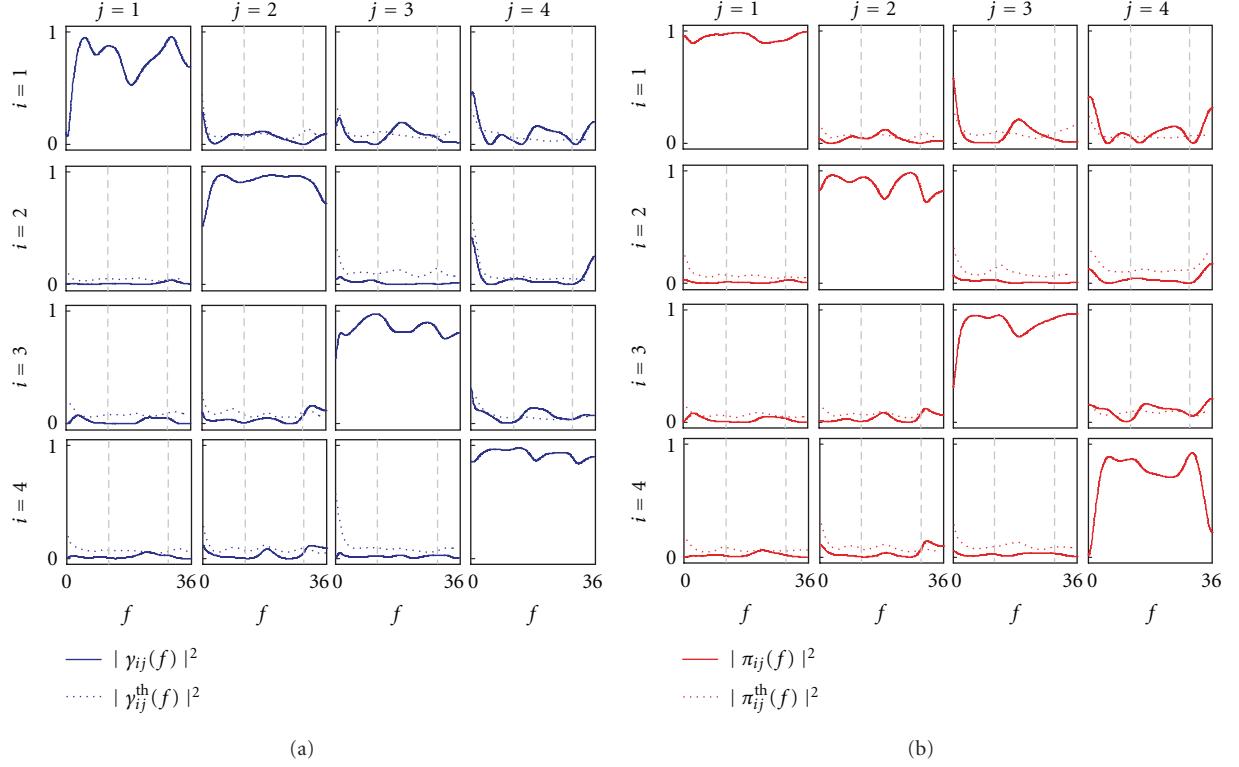


FIGURE 8: Frequency-domain measures of causality for the exemplary neurophysiological application (y_1 : left central EEG; y_2 : right central EEG; y_3 : parietal EEG; y_4 : occipital EEG). (a) Squared DC from y_j to y_i , ($|y_{ij}(f)|^2$) plotted together with its corresponding threshold for significance ($|y_{ij}^{\text{th}}(f)|^2$). (b) Squared PDC from y_j to y_i , ($|\pi_{ij}(f)|^2$) plotted together with its corresponding threshold for significance ($|\pi_{ij}^{\text{th}}(f)|^2$). Vertical dashed lines in each off-diagonal plot denote the bounds of the beta frequency band (13–30 Hz).

fully described by the autoregression so that, for example, important power amounts in specific frequency bands could not be properly quantified. When the whiteness test is not passed, the experimenter should consider moving to different model structures, such as MV dynamic adjustment forms having the general structure of MVAR networks fed by individual colored AR noises at the level of each signal [37, 38].

Failure of fulfilling the requirement of mutual independence of the residuals corresponds to violating the assumption of strict causality of the MVAR process. This is an indication of the presence of significant instantaneous causality, and occurs anytime the time resolution of the measurements is lower than the time scale of the lagged causal influences occurring among the observed series. This situation is common in the analysis of neural data, such as fMRI where the slow dynamics of the available signals make rapid causal influences appearing as instantaneous, or EEG/MEG where instantaneous effects are likely related to signal cross-talk due to volume conduction [56]. In this case, the spectral decompositions leading to the definition of DC and PDC do not hold anymore, and this may lead to the estimation of erroneous frequency-domain connectivity patterns like spurious nonzero DC and PDC profiles indicating causality or direct causality for connections that are actually absent [29, 63]. A possible solution to this problem is to

use a higher sampling rate, but this would increase the data size and—most important—would introduce redundancy that might hamper model identification [64]. A recently proposed approach is to incorporate zero-lag interactions in the MVAR model, so that both instantaneous and lagged effects are described in terms of the model coefficients and may be described in the frequency domain [29, 63, 65]. This approach is very promising but introduces non-trivial identification issues which could limit its practical utilization. In fact, ordinary least-squares identification, though recently proposed for identification of the extended model [65], is not feasible because it forces arbitrarily the solution; to guarantee identifiability without prior constraints, additional assumptions such as non-gaussianity of the residuals have to be posed [63, 66].

5.2. Model Order Selection. Even when the most suitable model structure is selected for describing the available MV dataset, model mis-specification may still occur as a consequence of an inappropriate selection of the model order. Model order selection is in fact an issue in real data analyses where the true order is usually unknown. In general, a too low model order would result in the inability to describe essential information about the MV process, while a too high-order would bring about overfitting effects

implying that noise is captured in the model together with the searched information. Therefore, a tradeoff needs to be reached between good data representation and reasonably low model complexity. While information criteria like AIC or BIC are very popular (see also Section 4.2), a correct model order assessment is rather difficult because the estimated order may not meet the user expectations (in terms of spectral resolution when it is too low, or in terms of interpretability of highly variable spectral profiles when it is too high), or may even remain undetermined as the AIC/BIC figures of merit do not reach a clear minimum within the range searched [25, 29, 67]. Simulation studies have shown that both underestimation and overestimation of the correct model order may have serious implications on connectivity analysis, with an increasing probability of missed and false-positive connections [68, 69]. A recent interesting result is the apparent asymmetry in the adverse effects on connectivity analysis of choosing a wrong model order, with more severe effects for underestimation than overestimation [69], which suggests to prefer higher orders while tuning this parameter in practical MVAR analysis. Adopted with the proper cautiousness, this choice would be good as it is also known to increase frequency resolution of connectivity estimates and to favor the achievement of whiteness for the model residuals.

5.3. Selection of Variables. The tools surveyed in this study to measure connectivity are fully multivariate, in the sense that they are based on MVAR analysis whereby all the considered time series (often more than two) are modeled simultaneously. This approach overcomes the known problems of repeated bivariate analysis applied to multiple time series, consisting, for example, in the detection of false coupling or causality between two series when they are both influenced by a third series [13, 70]. Nevertheless, in practical experimental data analysis, it is often not possible to have access to the complete set of variables which are relevant to the description of the physiological phenomena of interest. This issue goes back to the requirement of completeness of information stated for causality analysis [40], and raises the problem that unmeasured latent variables—often called unobserved confounders—can lead to detection of apparent connectivity patterns that are actually spurious, even when multivariate tools are at hand. Dealing with latent variables seems a daunting challenge, because there is no unique way to determine the information set relevant for a given problem. However, recent developments have started giving a response to this challenge through the proposition of approaches to causality analysis based on the idea that latent variables may give rise to zero-lag correlations between the available modeled series, and thus can be uncovered, at least in part, by further analyzing such a correlation [71].

A different but related problem to that of completeness is the redundancy in the group of the selected variables. Historically, the issue of redundant variables has been viewed more as a problem of increased model complexity and related decrease of parameter estimation accuracy in the modeling of massively MV data sets (such as those

commonly recorded in fMRI or high resolution EEG-MEG studies). This problem has been tackled through the introduction of network reduction approaches (e.g., [72]) or sparse (regularized/penalized) regression techniques (e.g., [73]), which allow to perform efficient high-dimensional MVAR analysis. More recent works have introduced a general formalism to recognize redundant variables in time series ensembles, showing that the presence of redundant variables affects standard connectivity analyses, for example, leading to underestimation of causalities [74, 75]. An elegant solution to this problem, proposed in [74], consists in performing a block-wise approach whereby redundancy is reduced grouping the variables in a way such that a properly defined measure of total causality is maximized.

5.4. Nonstationarity and Nonlinearity. The MVAR-based framework for connectivity analysis is grounded on the basic requirement that the set of observed multiple time series is suitably described as a realization of a vector stochastic process which is both linear and stationary. Despite this, nonlinear and nonstationary phenomena are abundant in physiological systems, and it is well known that MVAR-model analysis performed on nonlinear and/or nonstationary data may lead to a range of erroneous results [76]. In general, nonlinear methods are necessary to perform a thorough evaluation of connectivity whenever nonlinear dynamics are expected to determine to a non-negligible extent the evolution over time of the investigated time series. Analogously, when nonstationary data are expected to reflect connectivity patterns exhibiting physiologically relevant changes over time, it makes sense to use time-varying methods for the detection of coupling or causality. In these situations, several nonlinear/nonstationary time series analysis approaches may be pursued. Nonlinear methods range from local linear MVAR models exploited to perform local nonlinear prediction [6, 77] to nonlinear kernels [5] and to model-free approaches based on information theory [7, 78], phase synchronization [79], and state-space interdependence analysis [80]. As to nonstationary analysis, one simple approach is to study short-time windows which may be taken as locally stationary [67], while more complex but potentially more efficient approaches include spectral factorization of wavelet transforms [81] and the combination of state space modeling and time-dependent MVAR coefficients [82].

On the other hand, linear time-invariant analyses like the MVAR-based approach presented in this study remains of great appeal for the study of physiological interactions, mainly because of their simplicity, well-grounded theoretical basis, and shorter demand for data length in practical analysis. The problem of non-stationarity may be dealt with following common practical solutions like, for example, looking for analysis windows in which the recorded signals are stable and satisfy stationarity tests, and filtering or differentiating the data if necessary (though this has to be done cautiously [83]). The problem of nonlinearity may be faced looking for experimental setups/conditions in which the system dynamics may be supposed as operating, at the

level of the recorded signals, according to linear mechanisms. In fact, based on the observation that nonlinear systems often display extensive linear regimes, many neuroscience studies have shown that the linear approximation may suffice for describing neurophysiological interactions, especially at a large-scale level (see, e.g., the reviews [56, 84]). Even in circumstances where nonlinear behaviors are manifest, such as simulated chaotic systems or real EEG activity during certain phases of epileptic seizure, linear techniques have been shown to work reasonably well for the detection of connectivity patterns [11, 85]. In addition, we remark that the large majority of nonlinear methods used in the MV analysis of neurophysiological signals do not provide specific frequency-domain information [2]. The existing bivariate nonlinear frequency-domain tools, such as cross-bispectrum and cross-bicoherence [86], are useful to characterize dependencies between oscillations occurring in different frequency bands. However, analysis of coupling and causality between iso-frequency rhythms observed in different signals is intrinsically linear, and this further supports utilization of the linear framework for this kind of analysis.

6. Conclusions

In this tutorial paper, we have illustrated the theoretical interpretation of the most common frequency-domain measures of connectivity which may be derived from the parametric representation of MV time series, that is, Coh, PCoh, DC, and PDC. We have shown that each of these four measures reflects in the frequency domain a specific time-domain definition of connectivity (see Table 1). In particular, while Coh and PCoh are symmetric measures reflecting the coupling between two processes, they can be decomposed into non-symmetric factors eliciting the directional information from one process to another, these factors being exactly the DC and the PDC. Moreover, PCoh and PDC measure direct connectivity between two processes, while Coh and DC account for both direct and indirect connections between two processes in the MV representation.

We have pointed out the existence of a dual description of the joint properties of an MVAR process such that Coh and DC on one side, and PCoh and PDC on the other side, may be derived from the spectral matrix describing the process and from its inverse, respectively. This duality relationship highlights advantages and disadvantages of the various connectivity measures. Being related to spectral densities, Coh and DC provide meaningful quantification of coupling and causality in terms of (normalized) power shared by the two considered processes; on the contrary, PCoh and PDC are derived through an analysis performed in the inverse spectral domain which cannot provide evident physical information for the absolute values of the resulting indexes. On the other side, the procedure of “partialization” implicit in the computation of the inverse spectral matrix lets PCoh and PDC elicit the structural information of the MV process, so that they reflect direct connections only; this ability is not shared by Coh and DC, which mix together direct and indirect transfer pathways and thus cannot

provide a straightforward representation of the connectivity structure of the process. Another interesting observation comes from the decomposition of Coh (or, dually, of PCoh) between two processes into DC (or PDC) terms involving a third process (10a) and (10b); these relationships indicate that spurious (direct) coupling may be detected when the two processes under analysis, though not being truly connected, are connected to another common process.

The picture emerging from these results provides suggestions for the utilization of the various connectivity measures in the analysis of MV processes. First, measures of causality should be preferred to measures of coupling, as the latter cannot provide directional information and may be confusing as they are sensitive to spurious connectivity. Second, both DC and PDC should be considered as causality measures because they complement each other in terms of advantages and drawbacks: DC measures causality in meaningful physical terms as power contributions, but cannot separate direct effects from indirect ones; PDC determines the correct interaction structure in terms of direct causal effects, but its absolute values lack of straightforward interpretability. As to recommendations for the practical analysis of real MV time series, we remark the importance of validation tests, which constitute important safeguards against drawing erroneous inferences consequently to model misspecification, and of assessing the significance of each estimated connectivity measure, which is fundamental to provide statistical validity to the estimated MV process structure. Taking all these aspects into account, we have shown the practical applicability of the presented frequency-domain connectivity measures in neurophysiology. In the reported example, the simultaneous computation of Coh, PCoh, DC, and PDC, and of their specific significance thresholds, from multiple EEG recorded during the execution of a combined visuomotor task led us to infer the existence of a specific network subserving sensorimotor integration. This network was characterized by a significant coupling between visual and motor cortical regions, which was particularized into significant causality from the occipital to the left central cortex, suggesting a driving role of the visual feedback on the EEG activity of the motor areas.

References

- [1] B. Horwitz, “The elusive concept of brain connectivity,” *NeuroImage*, vol. 19, no. 2, pp. 466–470, 2003.
- [2] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, “Nonlinear multivariate analysis of neurophysiological signals,” *Progress in Neurobiology*, vol. 77, no. 1-2, pp. 1–37, 2005.
- [3] K. Friston, “Functional and effective connectivity: a review,” *Brain Connectivity*, vol. 1, no. 1, pp. 13–36, 2011.
- [4] B. Gourevitch, R. L. Bouquin-Jeannes, and G. Faucon, “Linear and nonlinear causality between signals: methods, examples and neurophysiological applications,” *Biological Cybernetics*, vol. 95, no. 4, pp. 349–369, 2006.
- [5] D. Marinazzo, W. Liao, H. Chen, and S. Stramaglia, “Nonlinear connectivity by Granger causality,” *NeuroImage*, vol. 58, no. 2, pp. 330–338, 2011.
- [6] L. Faes, A. Porta, and G. Nollo, “Mutual nonlinear prediction as a tool to evaluate coupling strength and directionality in

- bivariate time series: comparison among different strategies based on k nearest neighbors," *Physical Review E*, vol. 78, no. 2, Article ID 026201, 2008.
- [7] L. Faes, G. Nollo, and A. Porta, "Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique," *Physical Review E*, vol. 83, no. 5, Article ID 051112, 2011.
 - [8] I. Vlachos and D. Kugiumtzis, "Nonuniform state-space reconstruction and coupling detection," *Physical Review E*, vol. 82, no. 1, Article ID 016207, 2010.
 - [9] D. Chicharro and R. G. Andrzejak, "Reliable detection of directional couplings using rank statistics," *Physical Review E*, vol. 80, no. 2, Article ID 026217, 2009.
 - [10] J. Bhattacharya and H. Petsche, "Phase synchrony analysis of EEG during music perception reveals changes in functional connectivity due to musical expertise," *Signal Processing*, vol. 85, no. 11, pp. 2161–2177, 2005.
 - [11] M. Winterhalder, B. Schelter, W. Hesse et al., "Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems," *Signal Processing*, vol. 85, no. 11, pp. 2137–2160, 2005.
 - [12] M. J. Kaminski and K. J. Blinowska, "A new method of the description of the information flow in the brain structures," *Biological Cybernetics*, vol. 65, no. 3, pp. 203–210, 1991.
 - [13] Y. Chen, S. L. Bressler, and M. Ding, "Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data," *Journal of Neuroscience Methods*, vol. 150, no. 2, pp. 228–237, 2006.
 - [14] P. L. Nunez, *Neocortical Dynamics and Human EEG Rhythms*, Oxford University Press, New York, NY, USA, 1995.
 - [15] S. M. Kay, *Modern Spectral Estimation. Theory & Application*, Prentice Hall, Englewood Cliffs, NJ, USA, 1988.
 - [16] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1986.
 - [17] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance," *Biological Cybernetics*, vol. 85, no. 2, pp. 145–157, 2001.
 - [18] L. A. Baccala, K. Sameshima, G. Ballester, A. C. Valle, and C. Timo-laria, "Studying the interaction between brain structures via directed coherence and granger causality," *Applied Signal Processing*, vol. 5, no. 1, pp. 40–48, 1998.
 - [19] L. A. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.
 - [20] L. A. Baccala, K. Sameshima, and D. Y. Takahashi, "Generalized partial directed coherence," in *Proceedings of the 15th International Conference on Digital Signal Processing (DSP'07)*, pp. 163–166, Cardiff, Wales, UK, July 2007.
 - [21] D. Y. Takahashi, L. A. Baccala, and K. Sameshima, "Information theoretic interpretation of frequency domain connectivity measures," *Biological Cybernetics*, vol. 103, no. 6, pp. 463–469, 2010.
 - [22] B. Schelter, J. Timmer, and M. Eichler, "Assessing the strength of directed influences among neural signals using renormalized partial directed coherence," *Journal of Neuroscience Methods*, vol. 179, no. 1, pp. 121–130, 2009.
 - [23] L. Astolfi, F. Cincotti, D. Mattia et al., "Estimation of the cortical connectivity patterns during the intention of limb movements," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 4, pp. 32–38, 2006.
 - [24] J. R. Sato, D. Y. Takahashi, S. M. Arcuri, K. Sameshima, P. A. Morettin, and L. A. Baccala, "Frequency domain connectivity identification: an application of partial directed coherence in fMRI," *Human Brain Mapping*, vol. 30, no. 2, pp. 452–461, 2009.
 - [25] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by granger causality," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9849–9854, 2004.
 - [26] F. Amtage, K. Henschel, B. Schelter et al., "High functional connectivity of tremor related subthalamic neurons in Parkinson's disease," *Clinical Neurophysiology*, vol. 120, no. 9, pp. 1755–1761, 2009.
 - [27] S. Erla, L. Faes, E. Tranquillini, D. Orrico, and G. Nollo, "Multivariate autoregressive model with instantaneous effects to improve brain connectivity estimation," *International Journal of Bioelectromagnetism*, vol. 11, no. 2, pp. 74–79, 2009.
 - [28] S. Erla, L. Faes, G. Nollo, C. Arfeller, C. Braun, and C. Papadelis, "Multivariate EEG spectral analysis elicits the functional link between motor and visual cortex during integrative sensorimotor tasks," *Biomedical Signal Processing and Control*, vol. 7, pp. 221–227, 2012.
 - [29] L. Faes and G. Nollo, "Extended causal modelling to assess partial directed coherence in multiple time series with significant instantaneous interactions," *Biological Cybernetics*, vol. 103, no. 5, pp. 387–400, 2010.
 - [30] M. Kaminski, K. Blinowska, and W. Szlenberger, "Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness," *Electroencephalography and Clinical Neurophysiology*, vol. 102, no. 3, pp. 216–227, 1997.
 - [31] M. G. Philiastides and P. Sajda, "Causal influences in the human brain during face discrimination: a short-window directed transfer function approach," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, part 2, pp. 2602–2605, 2006.
 - [32] S. Gigi and A. K. Tangirala, "Quantitative analysis of directional strengths in jointly stationary linear multivariate processes," *Biological Cybernetics*, vol. 103, no. 2, pp. 119–133, 2010.
 - [33] L. Astolfi, F. Cincotti, D. Mattia et al., "Comparison of different cortical connectivity estimators for high-resolution EEG recordings," *Human Brain Mapping*, vol. 28, no. 2, pp. 143–157, 2007.
 - [34] Z. Albo, G. V. Di Prisco, Y. Chen et al., "Is partial coherence a viable technique for identifying generators of neural oscillations?" *Biological Cybernetics*, vol. 90, no. 5, pp. 318–326, 2004.
 - [35] L. A. Baccala and K. Sameshima, "Comments on 'Is partial coherence a viable technique for identifying generators of neural oscillations?': why the term 'Gersch causality' is inappropriate: common neural structure inference pitfalls," *Biological Cybernetics*, vol. 95, no. 2, pp. 135–141, 2006.
 - [36] M. Eichler, "On the evaluation of information flow in multivariate systems by the directed transfer function," *Biological Cybernetics*, vol. 94, no. 6, pp. 469–482, 2006.
 - [37] I. Korhonen, L. Mainardi, P. Loula, G. Carrault, G. Baselli, and A. Bianchi, "Linear multivariate models for physiological signal analysis: theory," *Computer Methods and Programs in Biomedicine*, vol. 51, no. 1-2, pp. 85–94, 1996.
 - [38] G. Baselli, A. Porta, M. Pagani, and S. Cerutti, "Spectral decomposition in multichannel recordings based on multivariate parametric identification," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 11, pp. 1092–1101, 1997.

- [39] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.
- [40] C. W. J. Granger, "Testing for causality: a personal viewpoint," *Journal of Economic Dynamics and Control*, vol. 2, pp. 329–352, 1980.
- [41] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, Wiley, Chichester, UK, 1990.
- [42] D. R. Brillinger, H. L. Bryant Jr., and J. P. Segundo, "Identification of synaptic interactions," *Biological Cybernetics*, vol. 22, no. 4, pp. 213–228, 1976.
- [43] R. Dahlhaus, "Graphical interaction models for multivariate time series," *Metrika*, vol. 51, no. 2, pp. 157–172, 2000.
- [44] M. Eichler, R. Dahlhaus, and J. Sandkuhler, "Partial correlation analysis for the identification of synaptic connections," *Biological Cybernetics*, vol. 89, no. 4, pp. 289–302, 2003.
- [45] M. R. Gevers and B. D. O. Anderson, "Representations of jointly stationary stochastic feedback processes," *International Journal of Control*, vol. 33, no. 5, pp. 777–809, 1981.
- [46] Y. Saito and H. Harashima, "Tracking of information within multichannel EEG record - causal analysis in EEG," in *Recent Advances in EEG and EMG Data Processing*, N. Yamaguchi and K. Fujisawa, Eds., pp. 133–146, Elsevier, Amsterdam, The Netherlands, 1981.
- [47] H. Lutkepohl, *Introduction to Multiple Time Series Analysis*, Springer, Berlin, Germany, 1993.
- [48] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [49] G. Schwartz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [50] A. Schlogl, "A comparison of multivariate autoregressive estimators," *Signal Processing*, vol. 86, no. 9, pp. 2426–2429, 2006.
- [51] M. Karimi, "Order selection criteria for vector autoregressive models," *Signal Processing*, vol. 91, no. 4, pp. 955–969, 2011.
- [52] S. Erla, L. Faes, and G. Nollo, "Robust estimation of partial directed coherence by the vector optimal parameter search algorithm," in *Proceedings of the 4th IEEE-EMBS Conference on Neural Engineering*, pp. 6280–6283, Antalya, Turkey, 2009.
- [53] L. H. Koopmans, *The Spectral Analysis of Time Series*, Academic Press, New York, NY, USA, 1974.
- [54] D. Y. Takahashi, L. A. Baccal, and K. Sameshima, "Connectivity inference between neural structures via partial directed coherence," *Journal of Applied Statistics*, vol. 34, no. 10, pp. 1259–1273, 2007.
- [55] B. Schelter, M. Winterhalder, M. Eichler et al., "Testing for directed influences among neural signals using partial directed coherence," *Journal of Neuroscience Methods*, vol. 152, no. 1-2, pp. 210–219, 2006.
- [56] S. L. Bressler and A. K. Seth, "Wiener-granger causality: a well established methodology," *Neuroimage*, vol. 58, no. 2, pp. 323–329, 2011.
- [57] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, pp. 346–382, 2000.
- [58] L. Faes, G. D. Pinna, A. Porta, R. Maestri, and G. Nollo, "Surrogate time series," *Surrogate Data Analysis for Assessing the Significance of the Coherence Function*, vol. 142, no. 7, pp. 1156–1166, 2004.
- [59] G. Nollo, L. Faes, A. Porta, R. Antolini, and F. Ravelli, "Exploring directionality in spontaneous heart period and systolic pressure variability interactions in humans: implications in the evaluation of baroreflex gain," *American Journal of Physiology*, vol. 288, no. 4, pp. H1777–H1785, 2005.
- [60] L. Faes, A. Porta, and G. Nollo, "Testing frequency-domain causality in multivariate time series," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, pp. 1897–1906, 2010.
- [61] J. M. Kilner, Y. Paulignan, and D. Boussaoud, "Functional connectivity during real vs imagined visuomotor tasks: an EEG study," *Neuroreport*, vol. 15, no. 4, pp. 637–642, 2004.
- [62] S. Erla, Computational methods for the assessment of brain connectivity in multisensory integration processes, Ph.D. thesis in Cognitive Neurosciences, University of Trento, Italy, 2011.
- [63] A. Hyvärinen, K. Zhang, S. Shimizu, and P. O. Hoyer, "Estimation of a structural vector autoregression model using non-gaussianity," *Journal of Machine Learning Research*, vol. 11, pp. 1709–1731, 2010.
- [64] E. K. Larsson, M. Mossberg, and T. Soderstrom, "An overview of important practical aspects of continuous-time ARMA system identification," *Circuits, Systems, and Signal Processing*, vol. 25, no. 1, pp. 17–46, 2006.
- [65] G. Deshpande, K. Sathian, and X. P. Hu, "Assessing and compensating for zero-lag correlation effects in time-lagged granger causality analysis of fMRI," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 6, pp. 1446–1456, 2010.
- [66] L. Faes, S. Erla, E. Tranquillini, D. Orrico, and G. Nollo, "An identifiable model to assess frequency-domain granger causality in the presence of significant instantaneous interactions," in *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'10)*, vol. 1, pp. 1699–1702, 2010.
- [67] M. Ding, S. L. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data pre-processing, model validation, and variability assessment," *Biological Cybernetics*, vol. 83, no. 1, pp. 35–45, 2000.
- [68] B. Schelter, M. Winterhalder, B. Hellwig, B. Guschlbauer, C. H. Lucking, and J. Timmer, "Direct or indirect? Graphical models for neural oscillators," *Journal of Physiology Paris*, vol. 99, no. 1, pp. 37–46, 2006.
- [69] E. Florin, J. Gross, J. Pfeifer, G. R. Fink, and L. Timmermann, "Reliability of multivariate causality measures for neural data," *Journal of Neuroscience Methods*, vol. 198, no. 2, pp. 344–358, 2011.
- [70] R. Kus, M. Kaminski, and K. J. Blinowska, "Determination of EEG activity propagation: pair-wise versus multichannel estimate," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 9, pp. 1501–1510, 2004.
- [71] S. X. Guo, A. K. Seth, K. M. Kendrick, C. Zhou, and J. Feng, "Partial Granger causality-eliminating exogenous inputs and latent variables," *Journal of Neuroscience Methods*, vol. 172, no. 1, pp. 79–93, 2008.
- [72] G. Deshpande, X. Hu, R. Stilla, and K. Sathian, "Effective connectivity during haptic perception: a study using Granger causality analysis of functional magnetic resonance imaging data," *NeuroImage*, vol. 40, no. 4, pp. 1807–1814, 2008.
- [73] P. A. Valdes-Sosa, J. M. Sanchez-Bornot, A. Lage-Castellanos et al., "Estimating brain functional connectivity with sparse multivariate autoregression," *Philosophical Transactions of the Royal Society B*, vol. 360, no. 1457, pp. 969–981, 2005.
- [74] L. Angelini, M. de Tommaso, L. Nitti, M. Pellicoro, and S. Stramaglia, "Redundant variables and Granger causality," *Physical Review E*, vol. 81, no. 3, part 2, Article ID 037201, 2010.
- [75] L. M. A. Bettencourt, V. Gintautas, and M. I. Ham, "Identification of functional information subgraphs in complex networks," *Physical Review Letters*, vol. 100, no. 23, Article ID 238701, 2008.

- [76] M. B. Priestley, *Non-Linear and Non-stationary Time Series Analysis*, Academic Press, London, UK, 1988.
- [77] W. A. Freiwald, P. Valdes, J. Bosch et al., "Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex," *Journal of Neuroscience Methods*, vol. 94, no. 1, pp. 105–119, 1999.
- [78] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy-a model-free measure of effective connectivity for the neurosciences," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 45–67, 2011.
- [79] M. G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, and R. Mrowka, "Identification of coupling direction: application to cardiorespiratory interaction," *Physical Review E*, vol. 65, no. 4, part 1, Article ID 041909, 2002.
- [80] A. Papana, D. Kugiumtzis, and P. G. Larsson, "Reducing the bias of causality measures," *Physical Review E*, vol. 83, no. 3, Article ID 036207, 2011.
- [81] M. Dhamala, G. Rangarajan, and M. Ding, "Estimating granger causality from fourier and wavelet transforms of time series data," *Physical Review Letters*, vol. 100, no. 1, Article ID 018701, 2008.
- [82] L. Sommerlade, M. Thiel, and B. Platt, "Inference of Granger causal time-dependent influences in noisy multivariate time series," *Journal of Neuroscience Methods*, vol. 203, no. 1, pp. 173–185, 2012.
- [83] E. Florin, J. Gross, J. Pfeifer, J. Pfeifer, G. R. Fink, and L. Timmermann, "The effect of filtering on Granger causality based multivariate causality measures," *Neuroimage*, vol. 50, no. 2, pp. 577–588, 2010.
- [84] M. Kaminski and H. Liang, "Causal influence: advances in neurosignal analysis," *Critical Reviews in Biomedical Engineering*, vol. 33, no. 4, pp. 347–430, 2005.
- [85] M. Jachan, K. Henschel, J. Nawrath, A. Schad, J. Timmer, and B. Schelter, "Inferring direct directed-information flow from multivariate nonlinear time series," *Physical Review E*, vol. 80, no. 1, part 1, Article ID 011138, 2009.
- [86] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectral Analysis: A Nonlinear Signal Processing Framework*, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.

Research Article

Statistical Analysis of Single-Trial Granger Causality Spectra

Andrea Brovelli

Institut de Neurosciences de la Timone (INT), UMR 7289 CNRS, Aix Marseille University, Campus de Santé Timone, 27 Bd. Jean Moulin, 13385 Marseille, France

Correspondence should be addressed to Andrea Brovelli, andrea.brovelli@univ-amu.fr

Received 18 November 2011; Revised 3 February 2012; Accepted 25 February 2012

Academic Editor: Luca Faes

Copyright © 2012 Andrea Brovelli. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Granger causality analysis is becoming central for the analysis of interactions between neural populations and oscillatory networks. However, it is currently unclear whether single-trial estimates of Granger causality spectra can be used reliably to assess directional influence. We addressed this issue by combining single-trial Granger causality spectra with statistical inference based on general linear models. The approach was assessed on synthetic and neurophysiological data. Synthetic bivariate data was generated using two autoregressive processes with unidirectional coupling. We simulated two hypothetical experimental conditions: the first mimicked a constant and unidirectional coupling, whereas the second modelled a linear increase in coupling across trials. The statistical analysis of single-trial Granger causality spectra, based on t -tests and linear regression, successfully recovered the underlying pattern of directional influence. In addition, we characterised the minimum number of trials and coupling strengths required for significant detection of directionality. Finally, we demonstrated the relevance for neurophysiology by analysing two local field potentials (LFPs) simultaneously recorded from the prefrontal and premotor cortices of a macaque monkey performing a conditional visuomotor task. Our results suggest that the combination of single-trial Granger causality spectra and statistical inference provides a valuable tool for the analysis of large-scale cortical networks and brain connectivity.

1. Introduction

The study of linear dependence between time series is central in many fields of science inferring causal relations among components of complex systems. The notion of causality between two time series was introduced by Wiener [1] and was later formalised by Granger within the framework of multivariate autoregressive (MVAR) linear models [2]. The definition of Granger-Wiener causality is based on statistical prediction: a time series has causal influence on another if the variance of the autoregressive prediction error of the later is reduced by including the past measurements of the former. Geweke demonstrated that pairwise time-domain Granger causality can be additively decomposed by frequencies [3] and introduced measures of directional linear dependence between two time series conditioned on a third [4]. Recently, Dhamala et al. [5] showed that Granger causality spectra can be estimated from Fourier and wavelet transforms of time series data, in addition to parametric spectral estimation methods.

In neuroscience, pairwise and conditional Granger causality spectra, based on parametric and nonparametric spectral methods, are becoming central for the analysis of interactions between neural populations within oscillatory brain networks [6–9]. In fact, current literature suggests that oscillations in neural populations activity, such as the local field potentials (LFPs), play a key role in modulating, filtering, and redirecting information in the nervous system [10–12]. Within this framework, Granger causality analysis is employed to reveal directional influences within oscillatory networks, such as during motor maintenance behaviours [13], and, more generally, it represents a crucial tool for the investigation of the neurophysiological substrate of cognitive functions [14].

Conventional research in neuroscience employs single-trial-experimental designs and performs statistical inference on single-trial dependent variables. The study of large-scale neural interactions and oscillatory activity, as measured with electroencephalography or magnetoencephalography (EEG/MEG) data and local field potentials (LFPs), is no

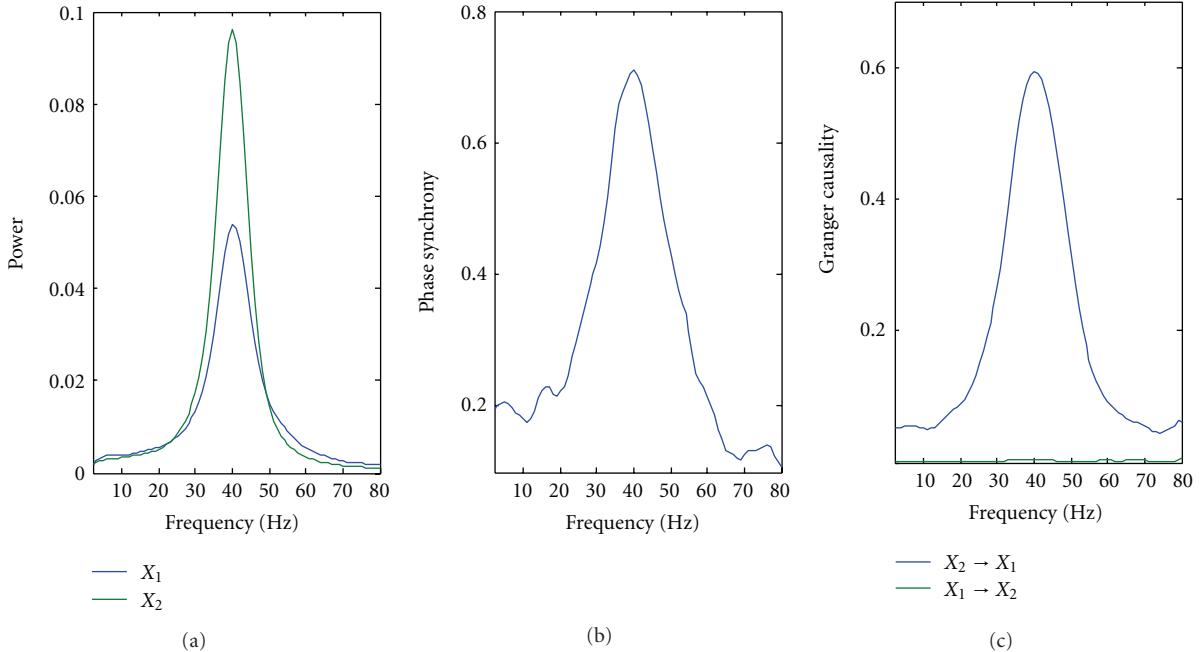


FIGURE 1: Simulations. Spectral analysis of synthetic data generated using a two-node network model with two autoregressive processes X_1 and X_2 and unidirectional coupling from X_2 to X_1 . (a) shows the power spectra for X_1 and X_2 , (b) shows their phase synchrony spectrum, and (c) depicts the Granger causality spectra.

exception. In fact, neural correlates of cognitive functions are classically searched in modulations of signal power and phase synchrony between channels. Current tools allow the estimation of spectral measures using Fourier and wavelet transforms on a single trial basis [15, 16]. Statistical inference can then be performed within the framework of general linear models (GLM) [17]. Non-parametric (or distribution-free) inferential statistical methods are also used when no assumption about the probability distributions of the dependent variables can be made. However, it is currently unclear whether single-trial estimates of Granger causality spectra, as can be computed using non-parametric methods [4, 5], can be used reliably to assess directional influences among neuronal oscillations. In the current paper, we addressed this issue by combining single-trial Granger causality spectra with statistical inference based on the GLM approach. We assessed the suitability of the approach through the analysis of synthetic data consisting of a two-node network model with two autoregressive processes. In addition, we tested the tools on one exemplar neurophysiological recording session consisting of a pair of LFPs recorded simultaneously from the dorsal premotor and lateral prefrontal cortex of a macaque monkey performing a conditional visuomotor task. Overall, the results suggest that the combination of single-trial Granger causality spectra and statistical inference provides a useful tool for the investigation of brain connectivity.

2. Materials and Methods

2.1. Synthetic and Neurophysiological Data. To investigate whether directional coupling between bivariate signals can be

inferred by combining single-trial Granger causality measures with parametric statistical tests, we analysed synthetic and neurophysiological data. Here is a description of the models used to generate synthetic data and the experimental methods of the neurophysiological recordings.

2.1.1. Synthetic Data. We considered a two-node network model with two autoregressive processes X_1 and X_2 and unidirectional coupling from X_2 to X_1 :

$$\begin{aligned} X_1(t) &= 0.35 X_1(t-1) - 0.5 X_1(t-2) + C X_2(t-1) + \epsilon_t, \\ X_2(t) &= 0.55 X_2(t-1) - 0.8 X_2(t-2) + \xi_t, \end{aligned} \quad (1)$$

where ϵ_t and ξ_t are Gaussian white noise processes with zero means and unit variances, C is the coupling strength. The sampling frequency was considered to be 200 Hz (similar to [5]). The signals display a peak at 40 Hz in power and phase synchrony spectra (Figures 1(a) and 1(b), the left and central panel, resp.). From the construction of the model, we can see that X_2 has a causal influence on X_1 (Equation (1) and Figure 1(c)). We performed three sets of simulations. In all simulations, each trial comprised 100 points (500 msec of simulated data for a sampling rate of 200 Hz). In the first set of simulations, we mimicked a hypothetical experimental condition with a constant and unidirectional (i.e., from X_2 to X_1) coupling strength $C = 0.3$. We generated a data set containing 100 sessions, each composed of 50 trials. This data set was then used to characterise the sensitivity of the statistical analysis with respect to trial number in each session. In other words, we studied the statistical power of the tests

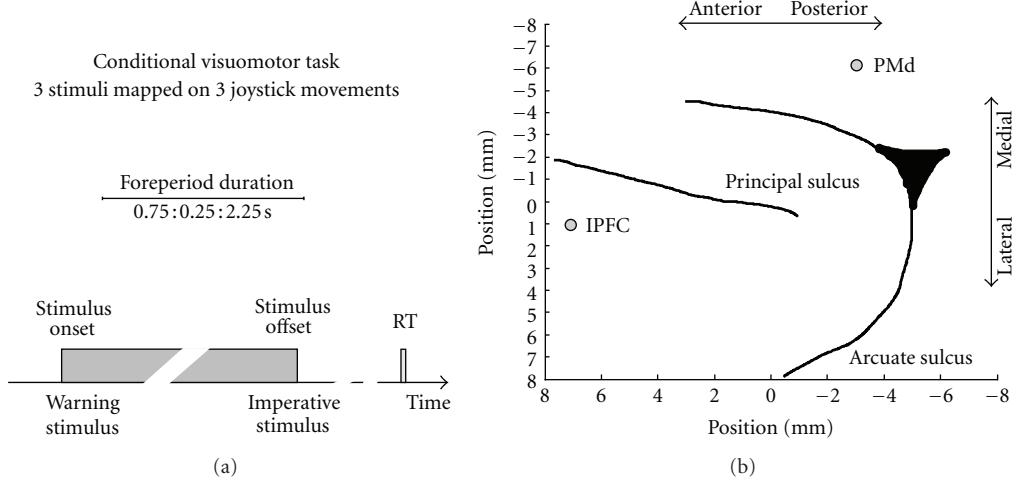


FIGURE 2: Neurophysiological experiments. (a) Conditional visuomotor task design. At each trial, the stimulus was presented at the centre of the screen (i.e., the warning stimulus) for a delay ranging from 0.75 to 2.25 seconds, in steps of 0.25 seconds (i.e., variable foreperiod paradigm). After stimulus offset (the instructive stimulus), the monkey had to execute the associated joystick movement to obtain reward. (b) Location of the two microelectrodes used to record the LFP analysed in the current paper. The electrodes were placed in the lateral prefrontal (IPFC) and dorsal premotor cortices (PMd).

by analysing data sets containing sessions with fewer trials (from 2 to 50 trials). In the second set of simulations, we generated synthetic data to investigate the range of coupling strengths that can be detected using the current approach. We generated 30 data sets (each containing 100 sessions and 50 trials) using coupling strengths C ranging from 0.01 to 0.3, in steps of 0.01 (30 possible values). In a third set of simulations, we modelled a linear increase in coupling across trials, as could be expected in experimental tasks exploring dynamic behaviours, such as during learning. We generated 100 sessions, each containing 150 trials, using coupling strength varying linearly across trials from 0 to 0.3 in steps of $0.3/m$, where m was the number of trials in each session. For example, in the simulations with 150 trials, the coupling strength C at trial 1 was equal to 0, it increased linearly at a rate of 0.002 every trial (i.e., $0.3/150$), and it was 0.3 at trial number 150. This dataset was also used to investigate statistical sensitivity as the number of trials in each session decreased from 150 to 4.

2.1.2. Neurophysiological Data. Neurophysiological recordings were conducted on a rhesus monkey at the Institut de Neurosciences Cognitives de la Méditerranée in the laboratory of Driss Boussaoud. Animal care and surgical procedures were in accordance with the European Communities Council Directive (86/609) for the use and care of laboratory animals in research. Results from literature suggest that the lateral prefrontal and dorsal premotor cortices play a key role in the acquisition and execution of arbitrary visuomotor associations (e.g., [18–20]). The aim of the electrophysiological study was to understand how these cortical areas coordinate during the acquisition and execution of arbitrary visuomotor associations. The entire neurophysiological database contains 93 recording sessions. In each session, the spiking activity of single neurons and the local

field potentials (LFPs) were recorded from up to 4 tungsten micro-electrodes simultaneously placed in the lateral pre-frontal (ventrolateral and dorsolateral prefrontal cortex, vLPFC and dLPFC, resp.) and dorsal premotor cortex (PMd). The analysis of the full dataset is beyond the scope of the current paper. However, the dataset represents an optimal benchmark to test our method on realistic neurophysiological bivariate data. Therefore, we analysed a pair of LFPs from one exemplar neurophysiological session.

The behavioural task required the monkey to perform a conditional visuomotor task associating three abstract images to three joystick movements (Figure 2(a)). The task design conforms a variable foreperiod (FP) paradigm. Stimulus onset can be defined as the warning stimulus (WS), and its stimulus disappearance represents the imperative stimulus (IS) instructing the monkey to perform the action. The fore-period duration (FP) is the time interval between the warning and imperative stimuli. A trial started when the animal held a joystick at a central position for 0.25 seconds. Thereafter, the stimulus was presented at the centre of the screen for a delay ranging from 0.75 to 2.25 seconds (the foreperiod duration), in steps of 0.25 seconds (7 possible delays). Delay durations were randomised across trials, and their offset instructed the monkey to execute the associated joystick movement, either to the right, up, or left. If movement direction was correct, a reward (fruit juice) was delivered after a fixed delay of 0.8 seconds; if incorrect, a purple circle appeared for 1.5 seconds as an error signal. The animal had 1 second to move the joystick in one of the three possible directions. If the response occurred late, the trial was aborted.

Local field potentials (LFPs) were simultaneously recorded from two electrodes (the sampling rate was 1000 Hz, and the raw signals were band-pass filtered from 1 to 250 Hz) placed in the lateral prefrontal cortex (IPFC) and dorsal

premotor area (PMd), respectively (Figure 2(b)). We analysed 259 epochs of 0.5 seconds in duration preceding the go cue (the imperative stimulus, IS) to present a single-case analysis illustrating how the neural oscillatory correlates of motor planning and/or expectation processes can be searched in the brain. To do so, we searched for trial-by-trial linear correlations between LFP spectral measures, such as power, phase synchrony, and Granger causality, and the IS expectancy. Imperative stimulus expectancy was estimated at each trial from the cumulative probability of IS occurrence, P_{IS} . Since the foreperiod delays ranged from 0.75 to 2.25 seconds (in steps of 0.25 seconds), the cumulative probability of IS occurrence (P_{IS}) was 1/7 for the shortest delay (0.75 seconds) and 1 for the longest (2.25 seconds). We then defined $S_{fp} = -\log(P_{IS})$ as the surprisal, or self-information, measuring the information content associated with IS occurrence. The surprisal in foreperiod duration S_{fp} is zero when the probability of occurrence of the go-signal P_{IS} is 1, that is, when the foreperiod duration is 2.25 seconds. In the analysis of the LFP, we correlated signal power, phase synchrony, and Granger causality at different frequencies using linear regression with the surprisal measure S_{fp} (more details in Section 2.3.3).

2.2. Single-Trial Granger Causality Spectra. To perform statistical inference based on parametric tests, we estimated Granger causality spectra on a single-trial basis. To do so, we computed the spectral density matrix for each trial using discrete fast Fourier transform (FFT) and Hanning window tapering of both synthetic and LFP time series data. The length of the FFT was 250 msec, stepped every 5 msec and zero-padded to 1 s to produce a frequency resolution of 1 Hz. Since each trial lasted 500 msec, the analysis produced 50 discrete FFTs for each trial. The spectra density matrix $S(n)_{lm}$ at trial n at channels l and m ($l, m = 1, 2$ in our case) was then given by

$$S(n)_{lm} = \langle X(\tau, n)_{l,n} \cdot X(\tau, n)_{m,n}^* \rangle_\tau, \quad (2)$$

where $X(\tau, n)_{l,n}$ is the FFT of a 250 msec epoch of signal centred at time lag τ within trial n , the expectation (denoted by $\langle \dots \rangle_\tau$) is taken over all FFTs at different time lags τ within a trial and $*$ denotes the complex conjugate. In other words, the “proper” ensemble averaging required to estimate the spectral matrix (e.g., equation 17.15 in [6]) is not performed over trials, but over time epochs within each trial. Note that we dropped the frequency suffix from the spectral measures for simplicity.

The single-trial spectra matrix $S(n)_{lm}$ was then factorised using Wilson’s algorithm to obtain the transfer function H and noise covariance matrix Σ [5, 8, 21, 22]. This step is critical in the estimation of Granger causality using non-parametric spectral analysis methods. The pairwise single-trial Granger causality spectra are then given by

$$I(n)_{l-m} = \frac{S(n)_{ll}}{S(n)_{ll} - (\Sigma_{mm} - \Sigma_{lm}^2/\Sigma_{ll}) |H(n)_{lm}|^2}. \quad (3)$$

2.3. Statistical Analysis of Single-Trial Granger Causality Spectra

2.3.1. General Linear Model Approach to Single-Trial Granger Causality Spectra. We adopted a general linear model (GLM) approach to analyse the single-trial Granger causality spectra. Granger causality measures issued from the synthetic and neurophysiological data are not normally distributed. Non-parametric statistical tests should be preferred. However, since that the GLM approach plays a key role in classical inference in neuroimaging and neurophysiology, we log-transformed Granger causality spectra to render the data approximately Gaussian (Figure 3) (refer also to [17]). The general linear model is normally expressed in matrix formulation,

$$Y = X\beta + e, \quad (4)$$

where Y is the dependent variable and is a column vector containing the data observations; e is a column vector of error terms; β is the column vector of model parameters ($\beta = [\beta_1, \dots, \beta_p]^T$, where p is the number of model parameters); X is $j \times p$ design matrix, whose column vectors represent the independent variables. Model parameters β were estimated using an ordinary least square method. In the current study, hypothesis testing and statistical inference were then performed using “contrast” vectors.

2.3.2. Analysis of Synthetic Data. We performed statistical analysis of single-trial Granger causality spectra computed from three simulations. The first data set was generated to simulate a constant and unidirectional coupling from X_2 to X_1 . Our goal was to assess whether the underlying pattern of directional influence could be recovered from the statistical analysis of the data, in particular through the use of paired two-sample t -tests. The t -test assessed whether the mean values of log-transformed Granger causality spectra from X_2 to X_1 at a given frequency were significantly greater than from X_1 to X_2 . Given that the synthetic data is generated using a unidirectional coupling from X_2 to X_1 , there is justification for testing for significant difference specifically in one direction only (one-sided t -test). According to the nomenclature used above, we let $Y(j) = [\log_{10} I(n)_{2 \rightarrow 1}, \log_{10} I(n)_{1 \rightarrow 2}]^T$ be a $2n \times 1$ column vector containing the two concatenated log-transformed single-trial Granger causality values at a given frequency and session (for simplicity, the suffices for frequency and session were dropped), where $j = 1, \dots, 2n$ and n is the number of trials. The two-sample t -test is built using a design matrix X with two columns and j rows with variables indicating group membership (ones and zeros). We tested the null hypothesis $H_0: \log_{10} I(n)_{2 \rightarrow 1} = \log_{10} I(n)_{1 \rightarrow 2}$ against the alternative hypothesis $H_1: \log_{10} I(n)_{2 \rightarrow 1} > \log_{10} I(n)_{1 \rightarrow 2}$ using the contrast $c = [1 - 1]^T$. The t -test was performed at each frequency from 2 to 80 Hz and for each session. This procedure leads to 100 t -values and associated P -values at each frequency. To characterise the sensitivity of the statistical analysis with respect to the number of trials analysed in each simulated session, we performed t -tests on the log-transformed Granger causality

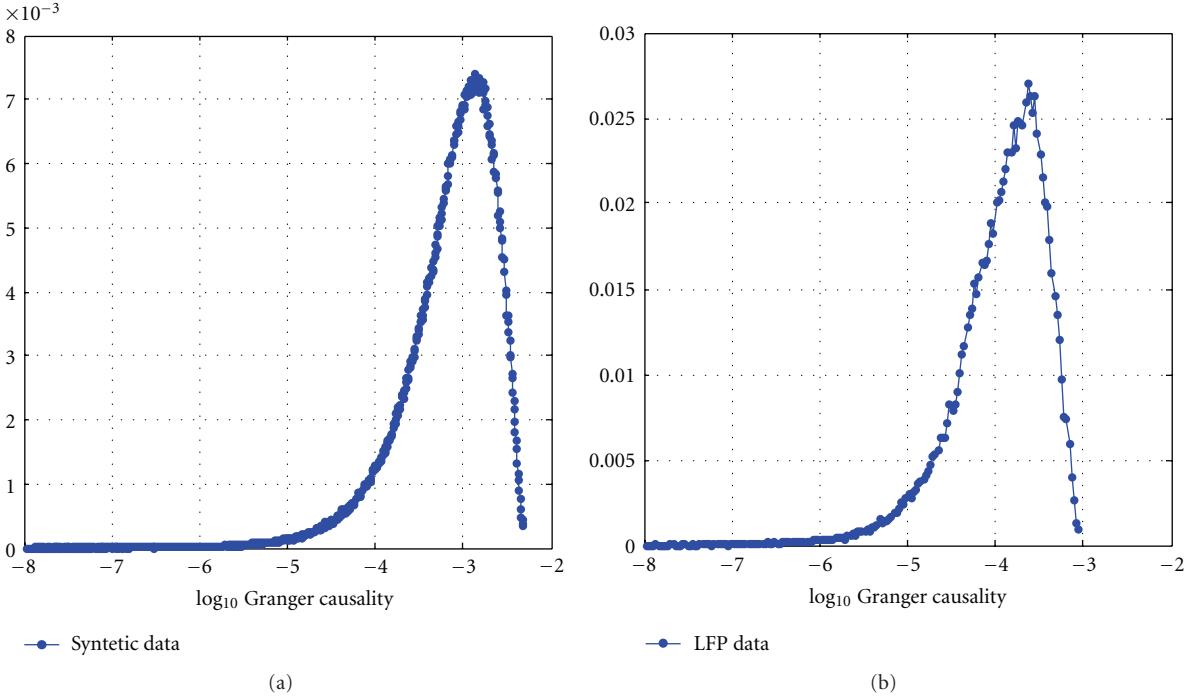


FIGURE 3: Distribution of log-transformed Granger causality data. Distribution of the log-transformation Granger causality values for synthetic (a) and LFP data (b). Histograms were computed using a number of bins equal to the square root of the number of elements in data.

values at 40 Hz (i.e., the peak frequency) and reduced the number of trials used in the statistical analysis, from 50 to 2 (i.e., 49 values). This produced 49 sets of 100 t - and P -values. Finally, we analysed the 49 sets of P -values to quantify the minimum number of trials required to detect unidirectional coupling from synthetic data. Assuming that the null hypothesis $H_0 \log_{10} I(n)_{2 \rightarrow 1} = \log_{10} I(n)_{1 \rightarrow 2}$ is false (i.e., the synthetic data was generated using a unidirectional coupling strength $C = 0.3$ from X_2 to X_1); we estimated the probability to perform type II errors for each trial number (from 2 to 50 trials) at three significance levels $\alpha = 0.01, 0.001$ and 0.0001 . As a reminder, type II errors occur when a false null hypothesis is accepted. The probability of committing this kind of errors is defined as β , and the power of the statistical test is $1-\beta$. By convention, we used a minimum required power of 0.8 as a cutoff for the determination of the minimum number of trials required to obtain a significant discrimination of the directional coupling in the data. In our analysis, type II error, occurred if the estimated P -value exceeded the significance level α . The probability β to perform Type II errors was estimated as the number of sessions displaying type II errors divided by the total number of sessions (100). The associated power was then given by $1-\beta$. This procedure was repeated for all trial numbers (from 2 to 50) so to give three curves of statistical power associated to three significance levels.

The second set of simulations was generated to assess the ability to detect smaller coupling strengths. We analysed 30 data sets (each containing 100 sessions and 50 trials)

each generated using different coupling strengths C varying from 0.01 to 0.3. We performed t -tests of the log-transformed Granger causality values at 40 Hz to test the null hypothesis $H_0 \log_{10} I(n)_{2 \rightarrow 1} = \log_{10} I(n)_{1 \rightarrow 2}$ against the alternative hypothesis $H_1 \log_{10} I(n)_{2 \rightarrow 1} > \log_{10} I(n)_{1 \rightarrow 2}$ using the contrast $c = [1 - 1]^T$. The t -test was repeated at all coupling strengths values to produce 30 sets of 100 t -values and associated P -values. To estimate the minimum coupling strength C detectable with the current approach, we performed power analysis (described in the previous paragraph) as a function of coupling strength at three levels of significance $\alpha = 0.01, 0.001$, and 0.0001 .

In the third set simulations, we modelled a linear increase in coupling strength C across trials. The simulations generated 100 sessions, each containing 150 trials. To retrieve the correct pattern of directional influence, we performed linear regression analysis. We let $Y(n) = [\log_{10} I(n)_{2 \rightarrow 1}]^T$ be a $n \times 1$ column vector containing the log-transformed single-trial Granger causality values at 40 Hz, where n is the number of trials. The design matrix X contained two columns and n rows. The first column modelled baseline and contained only ones, whereas the second column contained the values of the actual coupling strengths C as they varied across trials. The linear regression was then tested using a contrast $c = [0 1]^T$. To assess the sensitivity of the statistical analysis with respect to the number of trials analysed in each session, we reduced the number of trials used in the statistical analysis, from 150 to 3 stepped every trial (i.e., 148 values). This produced 148 sets of 100 t and P -values. The minimum number of trials

required to detect significant effects was estimated using the power analysis at three significance levels $\alpha = 0.01, 0.001$, and 0.0001 .

2.3.3. Analysis of Neurophysiological Data. We studied the relevance for neurophysiology by analysing two local field potentials (LFPs) simultaneously recorded from the prefrontal and premotor cortices of a macaque monkey performing a conditional visuomotor task. As mentioned in Section 2.1.2, the monkey performed a conditional visuomotor task based on a variable foreperiod (FP) paradigm. We searched for linear correlations between the surprisal in foreperiod duration S_{fp} (i.e., $-\log(P_{IS})$) and modulations in signal power, phase synchrony, and Granger causality at each frequency. Spectral analysis of LFP data was performed using the same parameters used for synthetic data. Briefly, the LFP power, phase synchrony, and Granger causality spectra were computed at single trials in a time window of 500 msec preceding the imperative stimulus (IS), using FFT sliding windows of 250 msec, stepped every 5 msec and zero-padded to 1 s. We performed linear regression analysis at each frequency between the trial-by-trial modulations in LFP spectral estimates (i.e., power, phase synchrony, and Granger causality) and the surprisal in foreperiod duration S_{fp} using the GLM approach described above. In other words, we let $Y(n)$ be an $n \times 1$ column vector containing the spectral measures (the log-transformed power, phase synchrony, Granger causality values), where n is the number of trials. The design matrix X contained two columns and n rows. The first column modelled baseline and contained only ones whereas the second column contained the values of the surprisal in foreperiod duration S_{fp} as they varied across trials. Since S_{fp} scales negatively with respect to the expectancy of IS occurrence, we tested a negative linear regression using a contrast $c = [0 \ -1]^T$. At each frequency, we obtained a t -value and associated P -value.

2.3.4. Software Implementation. All simulations and analyses were performed on MATLAB software (The MathWorks, Inc.). Spectral analysis was performed using functions from EEGlab software (<http://sccn.ucsd.edu/eeglab/>), whereas spectral matrix decomposition was performed using functions implemented in FieldTrip software (<http://fieldtrip.fcdonders.nl/>). Statistical analyses were performed using Matlab Statistical Toolbox.

3. Results and Discussion

3.1. Synthetic Data. We simulated a two-node network model with two autoregressive processes X_1 and X_2 , and unidirectional coupling from X_2 to X_1 (1). The sampling frequency was considered to be 200 Hz, leading to signals with a peak at 40 Hz in the power and phase synchrony spectra (Figure 1(a) and 1(b), resp.). From the construction of the model, X_2 had a linear causal influence on X_1 (Equation (1) and Figure 1(c)). To investigate whether directional influence between bivariate signals could be inferred by combining single-trial Granger causality measures

with statistical inference methods, we performed three simulations. In the first set of simulations, we mimicked a hypothetical experimental condition with a constant coupling across trials (i.e., $C = 0.3$). We then performed paired two-sample t -tests between the log-transformed Granger causality spectra to assess whether the mean values of log-transformed Granger causality spectra from X_2 to X_1 at a given frequency were significantly greater than from X_1 to X_2 . The t -test was performed at each frequency from 2 to 80 Hz and for each session (100 in total). This leads to 100 t -values and associated P -values at each frequency (see Section 2.3.2). Figure 4(a) depicts the boxplot representation of the distribution of the 100 P -values at each frequency. The results show that the statistical analysis is able to correctly infer the directional influence from X_2 to X_1 . In fact, the boxplots peak at 40 Hz and the P -values at 40 Hz are less than 10^{-6} (i.e., highly significant). To characterise the sensitivity of the statistical analysis with respect to the number of trials analysed in each session, we performed t -tests on the log-transformed Granger causality values at 40 Hz (i.e., the peak frequency) and reduced the number of trials used in the statistical analysis, from 50 to 2. Figure 4(b) shows the boxplot of the P -values over the number of trials simulated in each session. As expected, the mean P -values increase as the number of trials is reduced. To determine the minimum number of trials required to detect unidirectional coupling from synthetic data, we performed statistical power analysis (see Section 2.3.2). Briefly, statistical power is defined as $1 - \beta$, where β is the probability to perform type II errors (i.e., acceptance of a false null hypothesis) for three significance levels $\alpha = 0.01, 0.001$, and 0.0001 . The minimum number of trials required is defined as the first trial number whose statistical power exceeds 0.8. Figure 4(c) shows the statistical power for three significance levels. The minimum number of trials required for a statistical significance of 0.01, 0.001, and 0.0001 was 8, 13 and 18, respectively. To conclude, the results from the first simulation suggest that the combination of statistical inference based on parametric tests, such as t test, with single-trial Granger causality spectra successfully recovers the underlying pattern of directional influence with a limited number of trials.

In the second simulations, we investigated the range of coupling strengths that can be detected using the current approach. We generated synthetic data using coupling strengths varying from 0.01 to 0.3. Figure 5(a) shows the boxplot of the P -values as a function of coupling strength C . The P -values increase as the coupling strengths are decreased. To estimate the minimum coupling strength C detectable from synthetic data, we performed statistical power analysis (as described in the methods section) for three levels of significance $\alpha = 0.01, 0.001$, and 0.0001 . Figure 5(b) shows the statistical power curves, and it shows that the minimum coupling strengths detectable at statistical significances of 0.01, 0.001, and 0.0001 were 0.09, 0.11, and 0.14, respectively. The results therefore pointed out the limitation of the current approach to relatively high values of coupling strengths for a limited number of trials (i.e., 50 as in this case). However, for coupling strength greater than 0.14, the results suggest that directionality can be estimated with high significance.

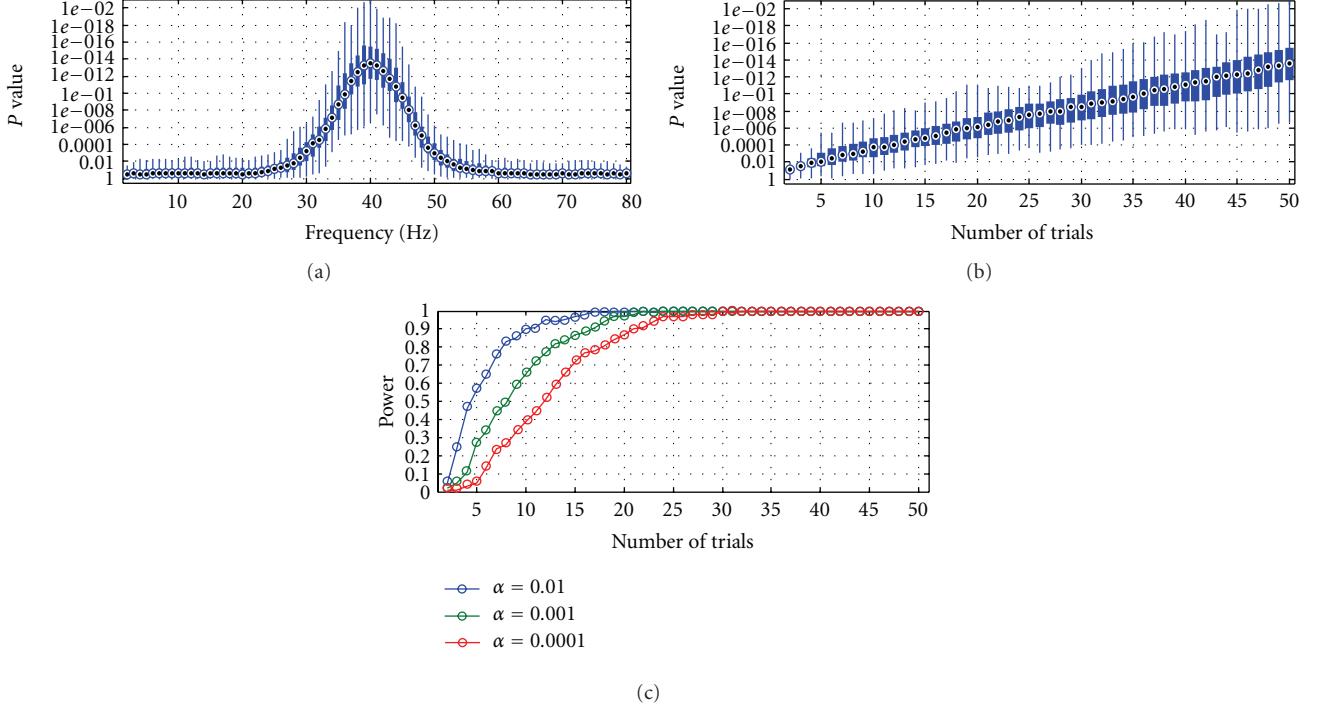


FIGURE 4: Simulation 1. (a) Granger causality p -spectra. Boxplot representation of the distribution of the P -values associated with each t -test. At each frequency, the circle in the box is the median value, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points. The P -values in each boxplot are associated with $H_1 \log_{10} I(n)_{2 \rightarrow 1} > \log_{10} I(n)_{1 \rightarrow 2}$. (b) Boxplot of the P -values plotted over the number of trials simulated in each session. (c) Statistical power curves at three levels of significance $\alpha = 0.01, 0.001$ and 0.0001 . By convention, the minimum required power of 0.8 was set as a cutoff for the determination of the minimum number of trials required to obtain a significant discrimination of the directional coupling in the data.

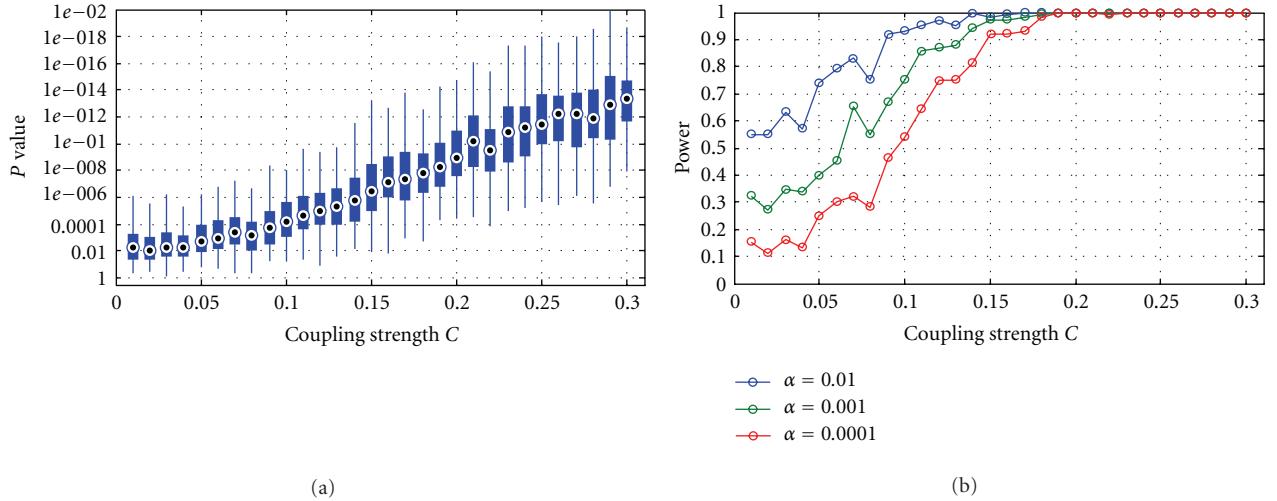


FIGURE 5: Simulation 2. (a) Granger causality p -spectra. Boxplot representation of the distribution of the P -values associated with each t -test as a function of the coupling strength C . (b) Statistical power curves at three levels of significance $\alpha = 0.01, 0.001$, and 0.0001 over the coupling strength C .

In the third simulations, we modelled a linear increase in directional influence from X_2 to X_1 varying from 0 to 0.3 across 150 trials (as could be expected in dynamic experimental tasks, such as during learning tasks). We performed linear regressions on a trial-by-trial basis between the

log-transformed Granger causality spectra from X_2 to X_1 and the coupling strengths. Figure 6(a) shows the Granger p -spectra displaying the boxplot representation of the P -values associated with the linear regressions performed at each frequency. The boxplots show that the statistical analysis based

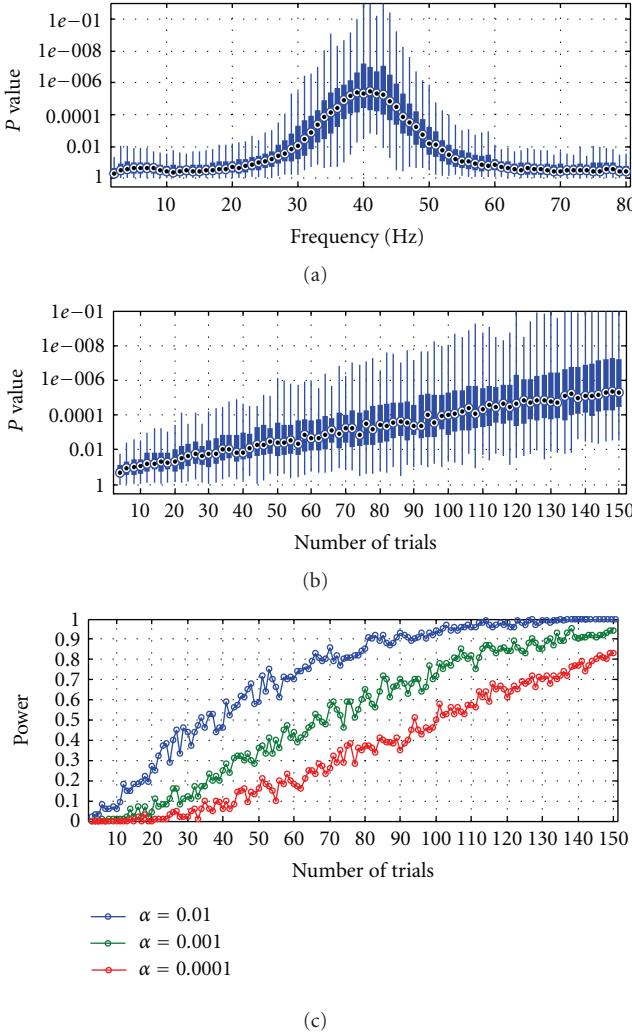


FIGURE 6: Simulation 3. (a) Granger causality p -spectra. Boxplot of the P -values associated with the linear regressions performed at each frequency between the log-transformed Granger causality spectra and the coupling strengths varying across trials. (b) Boxplot of the P -values plotted over the number of trials simulated in each dataset. (c) Statistical power curves at three levels of significance $\alpha = 0.01, 0.001$, and 0.0001 as a function of the number of trials used in each session.

on linear regression recovers the expected pattern of directional influence from X_2 to X_1 . Statistical P -value less than 0.01 are observed at the peak frequency of 40 Hz. To better quantify the minimum number of trials in each session required to significantly detect directional influence from the data, we performed linear regressions on the log-transformed Granger causality measures at 40 Hz (i.e., the peak frequency) on datasets containing fewer and fewer trials (from 150 to 4 trials). Figure 6(b) shows the boxplot of the P -values over the number of trials simulated in each dataset. Statistical power analysis showed that the minimum number of trials required for a statistical significance of 0.01, 0.001, and 0.0001 were approximately 70, 110 and 145, respectively (Figure 6(c)). Therefore, the results suggest that to detect

linear increases in coupling strengths across trials using linear regression, the minimum number of trials is larger than the number of trials required to detect constant directionality (Figure 4(c)).

The results of the simulations showed that single-trial estimates of Granger causality spectra can be used to quantify directional influence between bivariate synthetic data when combined with statistical inferences based on the GLM approach, such as t -tests (Figures 4(a) and 5(a)) and linear regression (Figure 6(a)). As expected, the method show sensitivity with respect to the number of trials used for statistical analysis (Figures 4(b) and 6(b)) and the coupling strength C (Figure 5(b)). The lower the number of trials and coupling strength, the less significant is the statistical analysis. The simulations quantified the minimum number of trials (Figure 4(c) and Figure 6(c)) and the range of coupling strengths (Figure 4(c)) required to detect directionality at different significance levels. To conclude, the simulation studies indicate that statistical inference based on general linear models (GLM) in combination with single-trial Granger causality spectra is a valuable tool to infer directional coupling among bivariate signals. Most importantly, the results suggest that the full range of statistical methods based on parametric (e.g., analysis of variance (ANOVA)) and non-parametric tests, general and generalized linear models, can be used to analyse single-trial Granger causality spectra issued from neurophysiological experiments. Even though the current work focused on Granger causality measures, we should stress that alternative methods can be used to quantify directional interactions, such as partial directed coherence (PDC), directed transfer function (DTF), directed transfer function (DTF), and transfer entropy (TE). While a detailed evaluation of the reliability of these measures was beyond the scope of the current paper (e.g., [23, 24]), we suggest that our simple statistical approach may be applied to alternative measures of directional coupling.

3.2. Neurophysiological Data. To investigate the feasibility of the current approach on realistic data, we analysed an exemplar neurophysiological session. Given that the monkey performed a conditional visuomotor task based on a variable foreperiod (FP) paradigm, we searched for linear correlations between the surprisal in foreperiod duration S_{fp} (i.e., $-\log(P_{IS})$) and modulations in signal power, phase synchrony, and Granger causality in a time window of 500 msec preceding the imperative stimulus (IS). At each frequency, we performed linear regression analysis between the trial-by-trial modulations in power, phase synchrony, and Granger causality with S_{fp} . Given that S_{fp} scales negatively with the probability of occurrence of the go cue (i.e., IS expectancy), we searched for negative linear correlations. Figure 7 shows the p -spectra associated with the linear regressions. Signal power in the lateral prefrontal cortex (blue curve Figure 7(a)) shows linear negative correlation with respect to S_{fp} , mainly in the beta range (from 15 to 30 Hz). A similar tendency is observed for the LFP in the PMd cortex, although not significant (green curve in Figure 7(a)). Phase synchrony

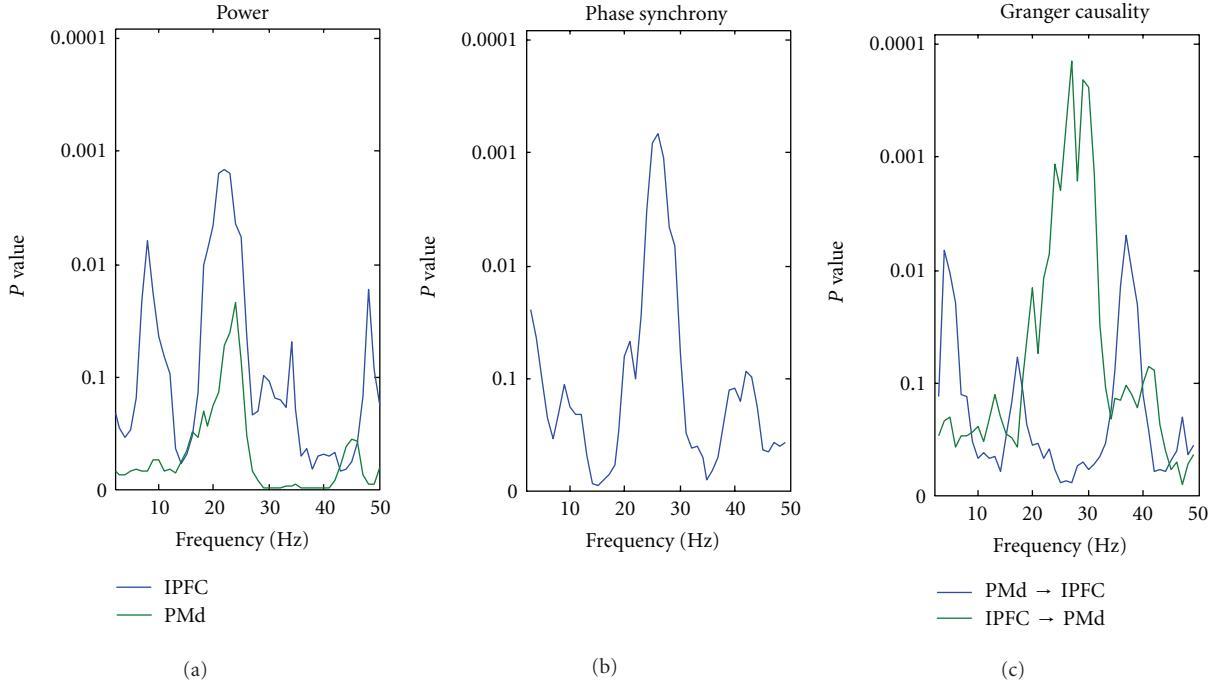


FIGURE 7: Neurophysiological results. Spectral analysis of neurophysiological data displaying p -spectra for (a) LFP power, (b) phase synchrony, and (c) Granger causality.

between the LFPs in the two cortical areas displays significant effects in the same frequency band, peaking around 25 Hz (Figure 7(b)). This indicates that the two areas oscillated synchronously in the beta range, and the degree of coherence scaled with the probability of occurrence of the go-cue. We then correlated the Granger causality measures with the values of S_{fp} , and we found that the coherence between the LFPs recorded in the two cortical areas can be explained by a unidirectional Granger causality influence from the lateral prefrontal cortex to the dorsal premotor area. The amount of Granger causality scales negatively with S_{fp} . This shows that the directional influence among the two areas increases with time as the probability of occurrence of the go cue increases. The prefrontal and premotor cortices, in addition to the basal ganglia, supplementary motor area, and cerebellum, have all been linked to the explicit estimation of duration [25]. Even though no conclusion can be drawn from the analysis of a single neurophysiological session, the current single-case analysis suggests a top-down effect of the lateral prefrontal cortex onto the dorsal premotor area. Further analyses of the full neurophysiological dataset are required to better understand the dynamic interplay between the prefrontal and premotor cortices in the prediction and update of temporal expectations as foreperiod unfolds. Overall, as an approach to large-scale cortical network analysis, our results suggest that statistical analyses of single-trial Granger causality spectra provides a valuable tool for in-depth investigation of the functional coupling of distributed neuronal assemblies.

4. Conclusions

The analysis of the synthetic data showed that directional coupling between bivariate signals can be inferred by combining single-trial Granger causality measures with parametric statistical tests based on a GLM approach. The statistical analysis of single-trial Granger causality spectra, based on t -tests and linear regression, successfully recovered the underlying pattern of directional influence. In addition, we characterised the minimum number of trials and coupling strengths required for significant detection of directionality. In fact, the number of trials required to obtain significant corresponds to conventional experimental situations. Finally, we demonstrated the relevance for neurophysiology by analysing two local field potentials (LFPs) simultaneously recorded from the prefrontal and premotor cortices of a macaque monkey performing a conditional visuomotor task. Our results suggest that the combination of single-trial Granger causality spectra and statistical inference provides a valuable tool for the analysis of large-scale cortical networks and brain connectivity. We suggest that the current approach may represent a simple statistical tool useful for the analysis of neurophysiological recordings issued from electroencephalographic (EEG), magnetoencephalographic (MEG), and intracranial EEG experiments. The approach may be extended to the full range of statistical methods based on parametric (e.g., analysis of variance (ANOVA)) and non-parametric tests, general and generalized linear models. Finally, we suggest that the same approach may be applied to alternative

measures of directional coupling based on the Granger causality principle.

Acknowledgments

The author would like to thank Dipesh Chaudhury who performed the neurophysiological recording of the LFP data analysed in the current paper under the supervision of Driss Boussaoud at the Institut de Neurosciences Cognitives de la Méditerranée (Marseille, France). In addition, the author would like to thank Salvatore Vaccarella from the International Agency for Research on Cancer (IARC, Lyon, France) for useful suggestions in statistical analysis.

References

- [1] N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers*, E. F. Beckenbach, Ed., vol. 1, McGraw-Hill, New York, NY, USA, 1956.
- [2] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.
- [3] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *Journal of the American Statistical Association*, vol. 77, pp. 304–313, 1982.
- [4] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *Journal of the American Statistical Association*, vol. 79, pp. 907–915, 1984.
- [5] M. Dhamala, G. Rangarajan, and M. Ding, "Estimating Granger causality from fourier and wavelet transforms of time series data," *Physical Review Letters*, vol. 100, no. 1, Article ID 018701, 4 pages, 2008.
- [6] M. Ding, Y. Chen, and S. L. Bressler, "Granger causality: basic theory and application to neuroscience," in *Handbook of Time Series Analysis*, M. Winterhalder, B. Schelter, and J. Timmer, Eds., pp. 437–460, Wiley-VCH, Berlin, Germany, 2006.
- [7] Y. Chen, S. L. Bressler, and M. Ding, "Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data," *Journal of Neuroscience Methods*, vol. 150, no. 2, pp. 228–237, 2006.
- [8] M. Dhamala, G. Rangarajan, and M. Ding, "Analyzing information flow in brain networks with nonparametric Granger causality," *Neuroimage*, vol. 41, no. 2, pp. 354–362, 2008.
- [9] M. le van Quyen and A. Bragin, "Analysis of dynamic brain oscillations: methodological advances," *Trends in Neurosciences*, vol. 30, no. 7, pp. 365–373, 2007.
- [10] P. Fries, "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence," *Trends in Cognitive Sciences*, vol. 9, no. 10, pp. 474–480, 2005.
- [11] P. Fries, "Neuronal gamma-band synchronization as a fundamental process in cortical computation," *Annual Review of Neuroscience*, vol. 32, pp. 209–224, 2009.
- [12] K. Benchenane, P. H. Tiesinga, and F. P. Battaglia, "Oscillations in the prefrontal cortex: a gateway to memory and attention," *Current Opinion in Neurobiology*, vol. 21, no. 3, pp. 475–485, 2011.
- [13] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9849–9854, 2004.
- [14] S. L. Bressler and V. Menon, "Large-scale brain networks in cognition: emerging methods and principles," *Trends in Cognitive Sciences*, vol. 14, no. 6, pp. 277–290, 2010.
- [15] C. Tallon-Baudry and O. Bertrand, "Oscillatory gamma activity in humans and its role in object representation," *Trends in Cognitive Sciences*, vol. 3, no. 4, pp. 151–162, 1999.
- [16] J. P. Lachaux, E. Rodrigues, M. le van Quyen, A. Lutz, J. Martinerie, and F. Varrela, "Studying single-trials of phase-synchronous activity in the brain," *International Journal of Bifurcation and Chaos*, vol. 10, no. 10, pp. 2429–2439, 2000.
- [17] S. J. Kiebel, C. Tallon-Baudry, and K. J. Friston, "Parametric analysis of oscillatory activity as measured with EEG/MEG," *Human Brain Mapping*, vol. 26, no. 3, pp. 170–177, 2005.
- [18] E. K. Miller, "The prefrontal cortex and cognitive control," *Nature Reviews Neuroscience*, vol. 1, no. 1, pp. 59–65, 2000.
- [19] S. P. Wise and E. A. Murray, "Arbitrary associations between antecedents and actions," *Trends in Neurosciences*, vol. 23, no. 6, pp. 271–276, 2000.
- [20] A. M. Graybiel, "Habits, rituals, and the evaluative brain," *Annual Review of Neuroscience*, vol. 31, pp. 359–387, 2008.
- [21] G. T. Wilson, "The factorization of matricial spectral densities," *SIAM Journal on Applied Mathematics*, vol. 23, no. 4, pp. 420–426, 1972.
- [22] G. T. Wilson, "A convergence theorem for spectral factorization," *Journal of Multivariate Analysis*, vol. 8, no. 2, pp. 222–232, 1978.
- [23] E. Florin, J. Gross, J. Pfeifer, G. R. Fink, and L. Timmermann, "Reliability of multivariate causality measures for neural data," *Journal of Neuroscience Methods*, vol. 198, no. 2, pp. 344–358, 2011.
- [24] D. Chicharro, "On the spectral formulation of Granger causality," *Biological Cybernetics*, vol. 105, no. 5–6, pp. 331–347, 2011.
- [25] J. T. Coull, R. K. Cheng, and W. H. Meck, "Neuroanatomical and neurochemical substrates of timing," *Neuropsychopharmacology*, vol. 36, no. 1, pp. 3–25, 2011.

Research Article

Enhancing the Signal of Corticomuscular Coherence

Cristiano Micheli^{1,2} and Christoph Braun^{2,3,4}

¹Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, 6525 EN Nijmegen, The Netherlands

²MEG-Center, University of Tübingen, Otfried-Müller-Straße 47, 72076 Tübingen, Germany

³Center for Mind/Brain Sciences (CIMeC), University of Trento, Via delle Regole 101, 38100 Trento, Italy

⁴Department of Cognitive and Educational Sciences (DiSCoF), University of Trento, Via delle Regole 101, 38100 Trento, Italy

Correspondence should be addressed to Cristiano Micheli, michelic72@gmail.com

Received 18 November 2011; Revised 7 February 2012; Accepted 20 February 2012

Academic Editor: Mingzhou Ding

Copyright © 2012 C. Micheli and C. Braun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The availability of multichannel neuroimaging techniques, such as MEG and EEG, provides us with detailed topographical information of the recorded magnetic and electric signals and therefore gives us a good overview on the concomitant signals generated in the brain. To assess the location and the temporal dynamics of neuronal sources with noninvasive recordings, reconstruction tools such as beamformers have been shown to be useful. In the current study, we are in particular interested in cortical motor control involved in the isometric contraction of finger muscles. To this end we are measuring the interaction between the dynamics of brain signals and the electrical activity of hand muscles. We were interested to find out whether in addition to the well-known correlated activity between contralateral primary motor cortex and the hand muscles, additional functional connections can be demonstrated. We adopted coherence as a functional index and propose a so-called nulling beamformer method which is computationally efficient and addresses the localization of multiple correlated sources. In simulations of corticomotor coherence, the proposed method was able to correctly localize secondary sources. The application of the approach on real electromyographic and magnetoencephalographic data collected during an isometric contraction and rest revealed an additional activity in the hemisphere ipsilateral to the hand involved in the task.

1. Introduction

One of the aims in applying recording techniques such as EEG and MEG, but also more recently ECoG, iEEG and laminar recordings, is to exploit the fine grained temporal resolution of the neuronal activity of the brain in order to quantify functional connectivity [1] between different brain areas and between the brain and external signals that might either be stimulation sequences or muscular activity in motor tasks.

Strong progress has been reported in the detection of brain areas revealing correlated activity by means of advanced source reconstruction techniques [2–8] such as beamformer. This methodology has successfully been applied in noninvasive EEG and MEG recordings and more recently also with invasive measurement of human brain activity [9]. Beamformers are capable of improving the signal-to-noise ratio (SNR) of the acquired brain activity

[10, 11], and thus enable a better reconstruction of the functional connectivity patterns among brain regions and between brain regions and external signals, than could be obtained by sensor level analysis.

In the present study, we focus on corticomotor connectivity reflecting peripheral motor control. To this aim, we acquired MEG data and the electromyogram (EMG) of finger muscles involved in a pinch grip task requiring isometric contraction. We used the oscillatory modulation of EMG activity to identify driving brain areas. As index for the synchronization between the external EMG signal and brain areas' specific activity we have chosen corticomuscular coherence (CMC, [12–15]).

Assuming a single driving brain area beamformer performs well as long as the signal-to-noise ratio is high enough. Yet, assuming multiple brain sources whose activity is correlated with the EMG signal, beamformers might fail to localize the respective brain regions because the

algorithm requires that the brain sources to be detected are linearly independent from each other [16]. However, in case of multiple sources being part of a larger network, this assumption might not be true and the beamformer approach will give faulty results.

Most importantly, if the muscular activity is modulated by the activity of more than one brain region at a time, a conventional beamformer will most likely localize only the most prominent modulatory source. It is well known from previous studies [5, 17] that the reconstruction of the brain sources correlated with an external source will result in the localization of the most correlated area and the cancellation of the other source. Given that it is not possible to localize multiple sources at a time, we identify the neural motor network in steps by progressively cancelling the most correlated sources with a technique called “nulling beamformer” [2, 6].

The present paper proposes a general pipeline which can be used to characterize cortico-motor connectivity in the source domain. We first adopt a DICS (dynamic imaging of coherent sources, [18]) beamformer and use coherence with the peripheral EMG signal as a metric to intercept the brain source showing strongest coherence. We then suppress this primary source by means of a nulling beamformer in order to uncover possible secondary (weaker) sources showing coherence with the EMG signal.

Additionally, we make use of a complementary step based on a subspace reprojection method [19, 20] to better identify both primary and secondary locations. This technique is necessary since the data is noisy and the localization can be improved by removing the local uncorrelated noise superimposed to the source signal. To demonstrate the validity of our approaches we applied them to simulated data of corticomotor coherence and to recordings in a real world experiment.

2. Methods

2.1. The DICS Beamformer with an External Reference Channel. The linearly constrained minimum norm (LCMV) beamformer is a spatial filter, which is used to separate the signal coming from a location of interest from other interferences. For a source, the time course of its activity can be expressed as the product of the raw data by the weights of the spatial filter:

$$\mathbf{d}(\mathbf{r}, t) = \mathbf{w}(\mathbf{r}) \mathbf{B}(t) \quad (1)$$

or, in the frequency domain:

$$\mathbf{d}(\mathbf{r}, f) = \mathbf{w}(\mathbf{r}, f) \mathbf{F}(f). \quad (2)$$

$\mathbf{B}(t)$ indicates the raw data, $\mathbf{F}(f)$ the Fourier transform of the data, $\mathbf{w}(\mathbf{r})$ and $\mathbf{w}(\mathbf{r}, f)$ the weights calculated with a time [16] or frequency domain [18] LCMV beamformer, and $\mathbf{d}(\mathbf{r}, t)$ and $\mathbf{d}(\mathbf{r}, f)$ the time course or the Fourier transform of the reconstructed dipolar source. The calculation of the spatial filter implies finding the multiplying weights that obey certain conditions.

Technically this is achieved by imposing that the variance of the source location is minimal, with the additional linear constraint that the signal originating from that location of interest (we call it the “virtual sensor”) is retained. This is equivalent to writing

$$\operatorname{argmin}_{\mathbf{w}} (\mathbf{w}^T \mathbf{C}_{\text{MEG}} \mathbf{w}), \quad \mathbf{w} \mathbf{L}(\mathbf{r}) = \mathbf{I}, \quad (3)$$

where \mathbf{I} indicates the unitary matrix, T is the matrix transpose operator, \mathbf{C}_{MEG} is the $N \times N$ covariance matrix of the MEG channels, and $\mathbf{L}(\mathbf{r})$ is the lead field for the location \mathbf{r} . By means of Lagrange function minimization the equation above leads to the following solution [5]:

$$\mathbf{w}(\mathbf{r}) = (\mathbf{L}(\mathbf{r})^T \cdot \mathbf{C}_{\text{MEG}}^{-1} \cdot \mathbf{L}(\mathbf{r}))^{-1} \mathbf{L}(\mathbf{r})^T \mathbf{C}_{\text{MEG}}^{-1}, \quad (4)$$

where -1 indicates the matrix inversion.

Hence the beamformer spatial filter is characterized by its weights: a set of N coefficients, being N the number of MEG channels. As a consequence all sensor level measures such as power, cross-spectral density, and coherence are translated into virtual sensors measures by a simple multiplication with the weights. As such

$$\mathbf{d}(\mathbf{r}, f) = \mathbf{w} \mathbf{F}(f) \quad (5)$$

is the Fourier transform of the source

$$P_{\text{MEG}}(\mathbf{r}) = \mathbf{w}^T \mathbf{F}^T \mathbf{F} \mathbf{w} \text{ or alternatively } P_{\text{MEG}}(\mathbf{r}) = \mathbf{w}^T \mathbf{C}_{\text{MEG}} \mathbf{w} \quad (6)$$

is the power of the source.

$$\text{csd}(\mathbf{r}) = \mathbf{w}^T \mathbf{C}_{\text{REF}} \quad (7)$$

is the cross-spectral density of the source, with \mathbf{C}_{REF} being the cross-spectral density between the MEG channels and an external source

$$\text{coh}(\mathbf{r}) = \frac{\text{csd}(\mathbf{r})^2}{\mathbf{P}_{\text{EMG}} \cdot \mathbf{P}_{\text{MEG}}(\mathbf{r})} \quad (8)$$

is the coherence between the source and the external signal and \mathbf{P}_{EMG} and $\mathbf{P}_{\text{MEG}}(\mathbf{r})$, are respectively, EMG and MEG sensors' power (at frequency f).

This particular fashion of beamformer is called reference-channel DICS and makes use of EMG-MEG coherence to localize brain synchronicity. It is important to note that the coherence parameter can be visualized as a 3D map in the head space of the subject and that DICS beamformer is frequency specific, as such the parameters in the previous equations are defined for one frequency bin at a time. The localization of the brain source which swings in synch with the EMG is operationalized by taking the maximum of the DICS coherence 3D map.

The coherence peak in the map, localized by means of DICS, takes the name of primary source or main source. To verify the reliability of DICS localization, we perform simulations with a priori known source locations and we

evaluate the significance of the coherence in subjects' data according to [15], as

$$\text{Significance} = 1 - (1 - \alpha)^{1/(L-1)}, \quad (9)$$

where α indicates the confidence level (95% in our case or $\alpha = 0.95$) and L the number of epochs.

2.2. The Nulling Beamformer. This technique is used to suppress the main activity, after its localization by means of the formerly described DICS technique. The implementation of the nulling beamformer begins with the construction of a modified lead field matrix $\tilde{\mathbf{L}}$, obtained by adding a term \mathbf{C}_S on the right side of the lead field matrix \mathbf{L} from the previous equations. This method requires the preselection of a ROI where the activity of the source has to be suppressed and defines the matrix \mathbf{C}_S as composed by the columns of the sources' lead fields to be cancelled. In formulas

$$\tilde{\mathbf{L}} = [\mathbf{L} \ \mathbf{C}_S], \quad (10)$$

$$\mathbf{w}(\mathbf{r}) = (\tilde{\mathbf{L}}(\mathbf{r})^T \cdot \mathbf{C}_{\text{MEG}}^{-1} \cdot \tilde{\mathbf{L}}(\mathbf{r}))^{-1} \tilde{\mathbf{L}}(\mathbf{r})^T \mathbf{C}_{\text{MEG}}^{-1}. \quad (11)$$

All quantities such as coherence and source power are defined as in the previous paragraph. Subsequently, a nulling constraint is imposed on the weights such as their spatial band pass and band stop characteristics are defined. This constraint is implemented by the multiplication with a coefficient c so that

$$\mathbf{w}_N(\mathbf{r}) = c\mathbf{w}(\mathbf{r}) \quad (12)$$

with $c = |1 \ 0 \ 0 \dots \ 0|$ for the scalar and $c = [I \ 0]$ for the vector beamformer, where unity values correspond to the band pass part of the filter and zeros to the band stop.

One of the drawbacks of this technique is that the degrees of freedom (defined as $M - 3 * J - 1$ ([21])) of the inversion term in (11) diminishes by a factor of $3 * J$ (M = number of channels, J = number of dipoles in the ROI [5]), making the aforementioned matrix close to singular. Therefore the choice of the ROI's radius is a trade-off between the extent of the area to be suppressed and the available degrees of freedom. A workaround to overcome the problem of insufficient degrees of freedom is to reduce the rank of the matrix \mathbf{C}_S by means of a singular value decomposition (as proposed in [6]). We normally apply the dimensionality reduction with a rejection percentage of 1 to 10% of the smallest eigenvalues.

The regional nulling beamformer is applied in this case to get rid of the primary cortico-muscular coherent activity to be able to visualize secondary sources. The attenuation on the unwanted primary activity in fact has the net effect of enhancing sources otherwise masked by the main localization, as demonstrated in Dalal's work [6].

2.2.1. Implementation Details. All analysis are run in Matlab (The Mathworks, Natick, MA, USA) and make use of FieldTrip, an open source toolbox for data analysis [22].

The nulling beamformer analysis is a novel and “nonstandard” part in the analysis pipeline (see Figure 2). Therefore its implementation is described in further details. All initial steps (referred to as “preprocessing”) are documented in the FieldTrip documentation pages: <http://fieldtrip.fcdonders.nl/tutorial/beamformer>. The names in italics that the reader encounters in this paragraph refer to specific FieldTrip instructions as they are typed in the Matlab environment.

The output variables of the previous analysis (see Figure 2, box “Inverse solution”) are (a) the volume conductor (describing the geometry of the head), (b) the positions of the sensors, (c) the trial-wise Fourier coefficients of the MEG channels, (d) the trial-wise Fourier coefficients of the EMG channel, (e) a “grid” structure containing (among others) the positions of the sources in 3D Cartesian coordinates and the corresponding lead fields describing the dipolar sources (forward solution), and (f) the coordinates of the point of maximal coherence as localized from DICS method. All reported quantities are stored as variables in the standard FieldTrip data structures (for a thorough reference on data structures please refer to the following page: http://fieldtrip.fcdonders.nl/faq/how_are_the_various_data_structures_defined).

Operatively the nulling beamformer involves the following steps:

- (1) definition of the ROI extent and construction of the \mathbf{C}_S matrix in (10);
- (2) modification of the leadfield matrices contained in the grid variable (the structure described previously at step (e)), by adding the columns of the \mathbf{C}_S matrix (or its reduced version) on the right side of each leadfield matrix;
- (3) calculation of the beamformer weights, as described in (11) (implemented by the function `ft_sourceanalysis`; the configuration options require `cfg.keepfilter = "yes"` and `cfg.method = "dics"` as input arguments for the function);
- (4) multiplication of the calculated weights by the matrix \mathbf{c} in (12), which defines the “band-pass” (=1) and “band stop” (=0) terms of each spatial filter;
- (5) calculation of the EMG-MEG cross-spectral density (csd) matrix for a defined frequency, using the function `ft_frequencyanalysis` (with the option `cfg.output = "csdandpower"`); note that the MEG and EMG datasets have to be previously appended (`ft_appenddata`);
- (6) creation of a `filter` field added to the grid structure (the previously mentioned FieldTrip structure, see step (e)) and containing the weights calculated in step 3;
- (7) projection of the previously calculated cross-spectral density (point 5) through to the weights. The projection step is implemented internally in the `ft_sourceanalysis` function and is accomplished by setting the arguments `cfg.method = "dics"` and `cfg.refchan = "EMGchannelname"`.

Note that each dipole's leadfield matrix has to be normalized by its Frobenius norm so that the beamformer localization is not affected by the depth bias [23]. One important issue regarding the nulling procedure is the inversion of the $(\tilde{\mathbf{L}}(\mathbf{r})^T \cdot \mathbf{C}_{MEG}^{-1} \cdot \tilde{\mathbf{L}}(\mathbf{r}))$ term in (11), which becomes rank deficient due to the modification of the leadfields [5]. To solve the problem the Moore-Penrose pseudoinverse [24] is usually applied (implemented in the Matlab function “*pinv*”; the pseudoinverse function used in FieldTrip, which is also called “*pinv*” is a subfunction of the “*beamformer_dics.m*” private function). In this version of *pinv* the tolerance is increased by a factor of 10 with respect to the standard *pinv* function). Most importantly the number of trials has to be superior or comparable to the number of MEG channels (a VSM-CTF system has 275, see “The Experiment” in paragraph 2.6.2) in order to estimate an unbiased channel level csd matrix (\mathbf{C}_{MEG} matrix in (11)).

2.3. The Data Subspace Reprojection. This technique has the aim of rejecting the noise in a ROI. The implementation consists in the definition of a geometrical region of interest and in the calculation of the Gram matrix related to the specific ROI, called Ω . The theoretical expression of the Gram matrix for a discrete points ROI Ω is:

$$\mathbf{G} = \sum_{r_i \in \Omega} \mathbf{L}(\mathbf{r}_i) \mathbf{L}^T(\mathbf{r}_i), \quad (13)$$

where $\mathbf{L}(\mathbf{r}_i)$ is the lead field defined for the voxel position \mathbf{r}_i .

Operatively, the algorithm requires to left multiply the data matrix $\mathbf{B}(t)$ by a matrix \mathbf{E} , whose columns are formed by the S larger eigenvectors of the Gram matrix ($S < N$: number of sensors). The selection criterion for the S is based on the variance explained by the first S sorted eigenvalues, according to:

$$\varepsilon = \frac{\sum_{i=1}^S \text{eig}(\mathbf{G})_i}{\sum_{i=1}^N \text{eig}(\mathbf{G})_i} \times 100. \quad (14)$$

In formula, we can call \mathbf{E}_S the matrix representing the S eigenvectors corresponding to the first S largest eigenvalues of \mathbf{G} , so that:

$$\mathbf{B}_{\text{DENOISE}}(t) = \mathbf{E}_S \mathbf{E}_S^T \mathbf{B}(t). \quad (15)$$

The described approach has been used in the context of EEG data analysis to get rid of spatial specific noise in a region of interest [19]. The consequence of the double matrix multiplication is equivalent to a PCA rejection of the small components and then a reprojection on the data space. This has the effect of reducing the noise due to the data outside of a certain ROI of interest and can be applied both to the main and to the secondary sources localizations in order to improve the signal to noise ratio of the local coherence.

2.4. The Complete Pipeline in a Flow Chart. A schematic representation of the processing which applies the proposed methods in a consistent pipeline is presented in Figure 2, where each block depicts a single step of the processing

flow. In particular the forward solution block implements the necessary steps to obtain the lead field matrices, used by the beamformer algorithm [5]. We make use of the Nolte solution [25], whereas more recent implementations of the MEG forward solution are available [26].

The preprocessing can be different according to the recordings and in the case of MEG implies the rejection of artifacts as described in [27]. Notably we make also use of independent component analysis (ICA) to accomplish this task [28]. No rectification is applied to the EMG since it has been recently objected to affect negatively the quality of the CMC analysis [29]. The spectral analysis makes use of multiple tapers to calculate both the spectral power and the cross-density matrix between MEG channels and the myography as described in [30]. We process epochs of 1 second with a number of 5 tapers.

The first step coming after frequency analysis is the DICS beamformer, which takes as inputs the spectra of both MEG and EMG signals, their cross spectrum and the lead fields resulting from forward model calculation. The visualization step generates a map of cortico-muscular coherence in three orthographic projections and highlights the maximum value (main activation). The region of interest (ROI) analysis implies the selection of the sources around the peak of activity in a radius of 3 cm from it. This step results in the extraction of the lead fields corresponding to the sources included in the ROI. These lead fields are used both for the nulling beamformer and for the beamspace reprojection method.

Successively the pipeline is composed by two additional steps: one responsible for the spatial band stop filtering of the data and the subsequent localization of the secondary CMC sources, the second responsible for the visualization of the results. The subspace reprojection technique is depicted in blue in Figure 2 and is generally applied to enhance the SNR of the localized sources.

2.5. The Statistical Analysis. The threshold of significance for the coherence maps was determined by a randomization. The first step of the statistical analysis shuffles the trials of the EMG's Fourier transform and generates a new set of complex coefficients at a certain frequency (the peak of sensor level coherency). For the specific goal of this paper we use $N_{\text{perm}} = 100$ permutations. Successively we calculate the beamformer coherence for all iterations, which results in as many vectors as numbers of permutations.

The P values for the randomization test (see paragraph 2.3 of Maris and Oostenveld paper [31]) are obtained from

$$P(i)_k = \frac{\sum_k \delta(\text{coh}(i)_k - \text{coh}(i))}{N_{\text{perm}}}, \quad (16)$$

where $i = 1 \dots N_{\text{vox}}$ is the voxel index, k is the permutation index ($k = 1 \dots N_{\text{perm}}$), $\text{coh}_k(i)$ is the randomized coherence for voxel i at permutation k , $\text{coh}(i)$ is the nonrandomized coherence calculated at voxel i and δ is a Kronecker function, which follows the rules:

$$\delta = 1 \text{ if } \text{coh}_k - \text{coh} > 0, \quad \delta = 0 \text{ if } \text{coh}_k - \text{coh} \leq 0. \quad (17)$$

The resulting P values which are smaller than a significance threshold for the statistical test (we choose 5%) are selected and correspond to spatial locations in the coherence map. The minimum coherence value of the extracted pool of voxels represents the empirical threshold.

2.6. Details of the Experiment and the Simulations

2.6.1. The Simulations. The simulation paradigm consists of two datasets: (1) a simulated MEG field containing two oscillating sources which are correlated with a simulated EMG signal, (2) a simulated MEG field containing one oscillatory source and a simulated EMG signal correlated with it. The first simulation is meant to show the effectiveness of the nulling beamformer in the suppression of a brain source and in the enhancement of the secondary one. In the second simulation, noise is added gradually in 10 different datasets, to show the degradation of the beamformer signal and the partial enhancement of the SNR due to the data reprojection technique using a ROI of 2 cm around the source and a number of $\varepsilon = 10^{-7}$ to select the largest eigenvalues.

The time courses of both simulation 1 and simulation 2 are generated by multiplying a Gaussian envelope with a 20 Hz sinusoidal signal in an epoch of 1 second (see Figure 1). The peak of the sinusoid is nominally fixed at 0.5 seconds and two jitters are defined across the 100 trials for phase and amplitude of the sinusoid. The coupling between and external signal and the simulated sources are defined by assigning a deterministic and a random component to the aforementioned jitters, according to the formula:

$$\begin{aligned} \text{MEG_jitter}_i &= [\text{det}_i * \text{CS} + \text{rand}_i * (1 - \text{CS})] * \text{nominal}, \\ \text{EMG_jitter}_i &= \text{det}_i * \text{nominal}, \end{aligned} \quad (18)$$

where MEG_jitter_i and EMG_jitter_i represent the jitter for MEG and MEG time courses at epoch i , CS is the coupling strength (a parameter defined between 0—no coupling—and 1—perfect coupling—), det_i is a realization of random noise which is common to EMG and MEG signals, rand_i is a realization of random noise that is present only in MEG signal, and nominal is the nominal value of the jitter parameter (phase or amplitude jitter).

The time courses are then multiplied for the lead fields of two sources defined in positions (0, -4, 10) and (0, 4, 10) cm in head Cartesian coordinates according to the CTF axes conventions [32]. The resulting simulated magnetic field is then added with random noise with varying intensity. In the first simulation the intensity varies from 10 to 100 times the root mean square value (rms) of the clean dataset, whereas in the second simulation the level of superimposed noise is 0.2 times the rms of the clean dataset.

The rms of a dataset is defined as

$$\text{rms} = \sqrt{\frac{\sum_i (x_i)^2}{N}}, \quad (19)$$

where x_i is a sample of the dataset and N is the total number of samples. The coupling strength for the first simulation is 1

for all jitters, whereas for the second simulation we choose 1 for the amplitude jitter and 0.9 for the phase jitter.

2.6.2. The Experiment. The experiment was run with a whole head, 275 channels MEG system (CTF/VSM Inc. Port Coquitlam, Canada) equipped with first-order axial gradiometers with 5 cm baseline and installed inside a magnetically shielded room (MSR, Vacuumschmelze Hanau, Germany). During the recording time the subject sat on a chair whose position and height could be regulated on demand. Its position was set such that the subject could see a feedback image projected on a screen of dimensions 42×32 cm (width \times height) situated at a distance of 0.4/0.5 m. An ad hoc built pinch device, which had a strain gauge force transducer mounted on it, was attached to the armrest of the chair in a position that was comfortably reachable by the subject's right hand. Five subjects (all aged between 30 and 35, all right handed, 4 males and 1 female) were selected for the experiment with the aim of carrying out a subject-specific analysis (no grand averages were planned for this experiment). A muscular activity (EMG) was recorded using bipolar derivations with electrodes of 11 mm diameter (In Vivo Metric, Healdsburg USA) mounted on the dorsal surface of first dorsal interosseous muscle at 20 mm distance from each other. Electrode gel "Abraleit light" (Falk Minow Services, Herrsching, Germany) was used to establish electrical contact between skin and electrodes with impedances below 20 kOhm. The EMG was digitized and stored in the same datasets as the MEG recordings. The subject had to grasp the pinch device and elicit a continuous constant force of 1 Newton for the whole duration of the experiment.

3. Results

3.1. The Simulations' Results. Simulation 1 shows that the nulling beamformer is able to suppress the left source and enhance the right one (Figure 3). Note that the dataset contains the time courses of both sources, but the classic beamformer approach could localize only the source with the highest SNR (see Figure 1). Due to this cancellation effect, it is impossible to proceed in the localization of putative weaker sources without recurring to alternative techniques as the one described here.

Simulation 2 shows the increase in SNR (Figure 3(b)) after the application of the subspace reprojeciton using the selected ROI. The upper picture shows the coherence profiles along the coronal slice (depicted in Figure 3(a)) and each line represents the DICS result for a different dataset. The results with a low SNR are in light blue and the ones with a high SNR in magenta according to the color scheme convention named "cool" in MATLAB. Accordingly, the lower picture shows the results for the same datasets after the combination of subspace reprojeciton and DICS. The SNR color scheme increases from black to yellow (color map "hot" in MATLAB). Note the general increase of SNR of the "hot" curves with respect to the "cool" ones, caused by

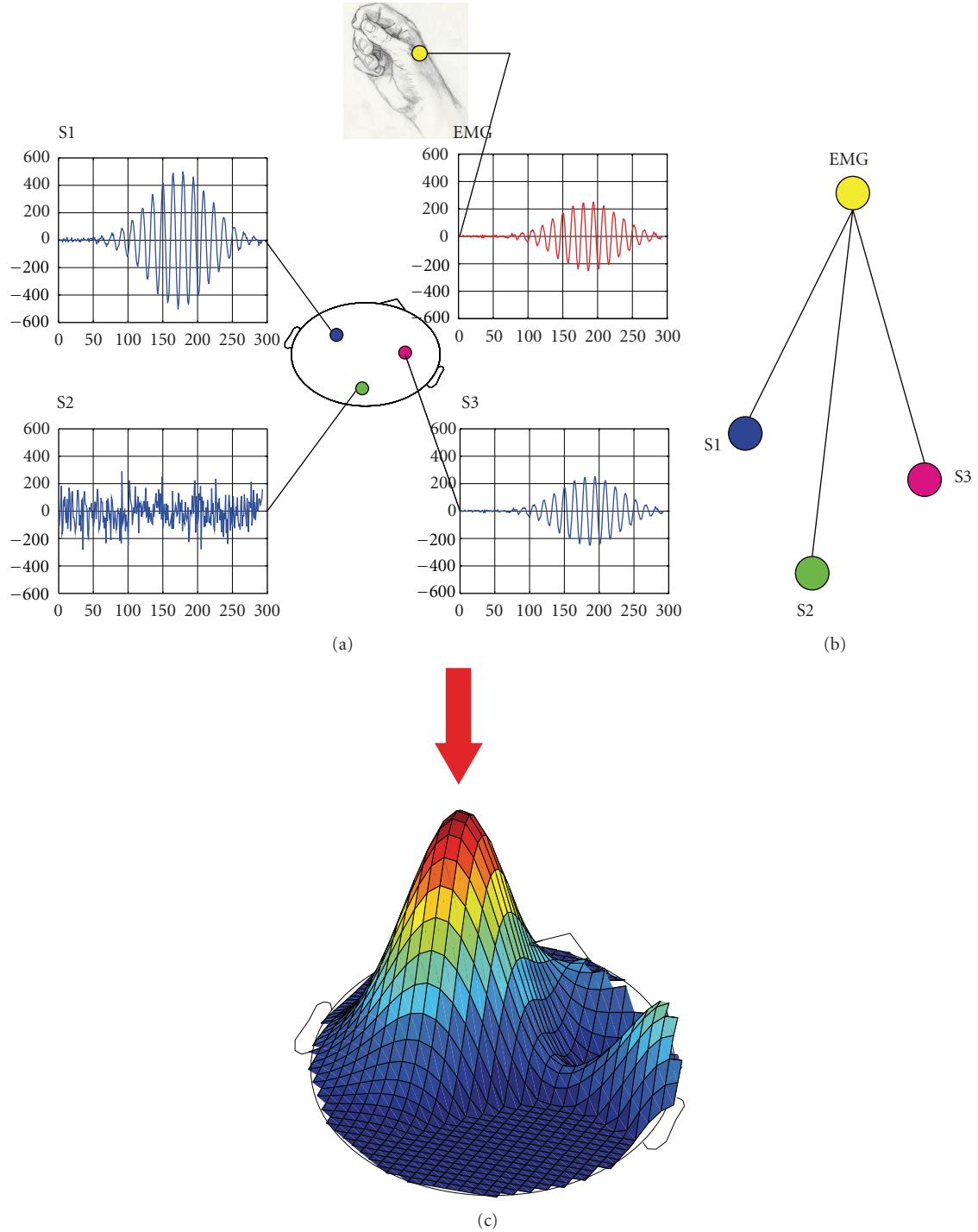


FIGURE 1: A representation of the motor network and of how the connectivity is calculated. (a) A representation of simulated brain sources in a volumetric slice (blue time courses, marked as S1, S2, and S3) and hand surface myography correlated with tracks S2 and S3 (EMG, red time course). (b) The scheme of functional connectivity. Coherence is calculated between EMG and all other sources. (c) A schematic output of DICS beamformer coherence. Note the presence of only one peak corresponding with the source having the highest SNR (S1). The red arrow indicates that this peak has to be suppressed by the nulling beamformer in order to localize the second EMG-correlated source (S3).

the noise reduction yielded by the aforementioned subspace reprojection technique.

3.2. The Subjects' Data. One of the five subjects (subject 5) did not show any muscular-MEG coherence at the sensor level and therefore was discarded from further analyses.

3.2.1. The Results of Beamspace Reprojection. One representative subject is selected to show the results of the applied technique (subject 1). Figure 4 shows the results of DICS beamformer before (upper row) and after (lower row) the application of the subspace reprojection technique on the preprocessed dataset. This illustrative example shows that the

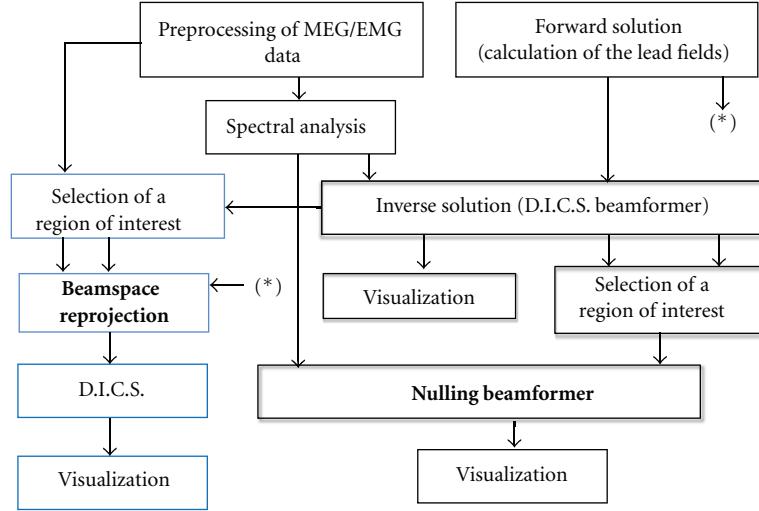


FIGURE 2: Scheme of the processing for beamformer localization of EMG coherent sources. The boxes indicate the methods described in this paragraph, with particular emphasis for the beamformer steps (shadowed boxes). The bold text in the boxes represents the algorithms contributed in the present paper. The light blue boxes refer to the beamspace reprojection pipeline, used to enhance the visualization of the main CMC source.

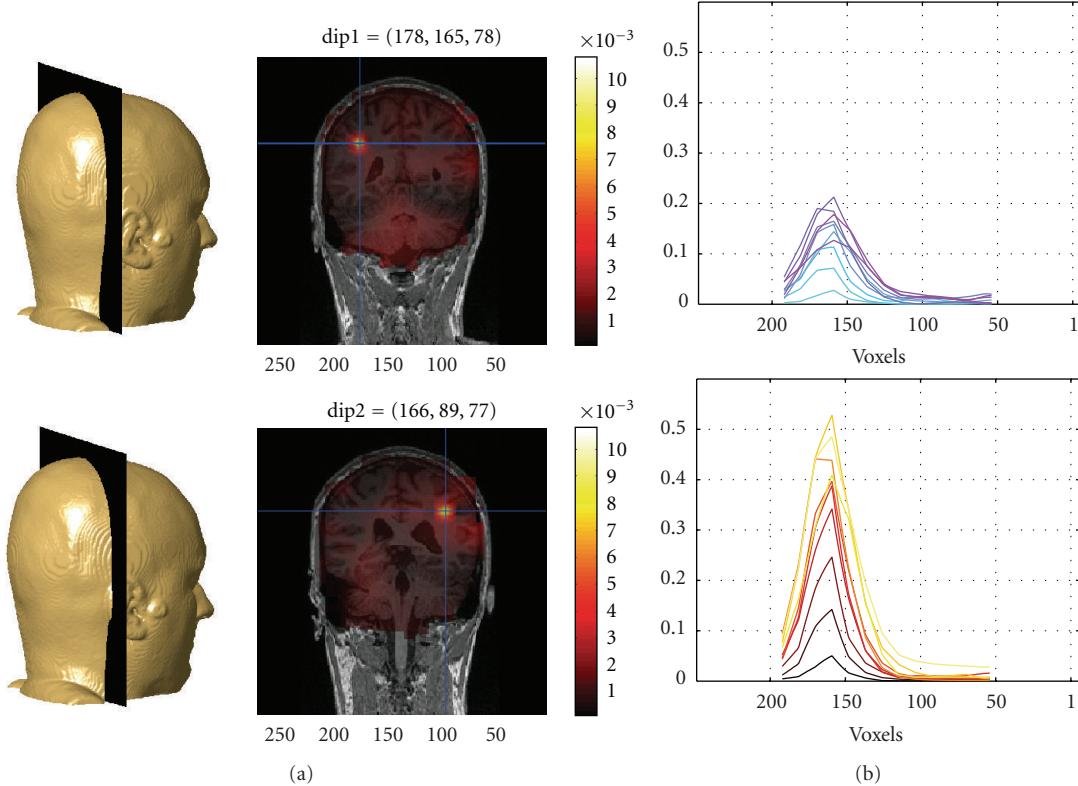


FIGURE 3: Results of the simulations. (a) The localization of two EMG-correlated sources is achieved in consecutive steps: with a standard DICS method (top-right panel, sagittal slice) and with DICS combined with regional suppression (bottom right panel). The leftmost source (dip1) is suppressed with a ratio of 30 (source coherence ratio), while the rightmost source is enhanced with a ratio of 17. (b) The localization of a single source with ten different SNR levels (cyan is the lowest and magenta is the highest) is achieved with DICS (top panel). The same simulated source+noise datasets are processed after local removal of noise by means of subspace reprojection (black is the lowest SNR, yellow is the highest).

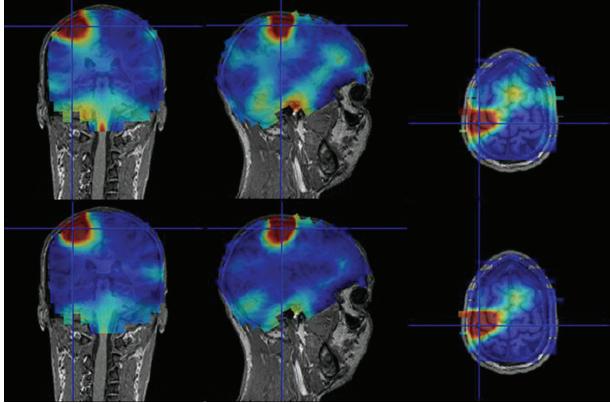


FIGURE 4: The subspace reprojection applied to subject 1. The three plots on top represent the coronal, sagittal, and transversal projections of the three-dimensional coherence map, before the application of beamspace reprojection. The three plots on the bottom depict the same projections of estimated activity, after the reduction of spatial noise. The technique is used to enhance the visualization of the coherence map. Both sets of plots refer to the same color scale (interval ranges from 0 to 0.008 of estimated coherence), and “jet” colormap is used (blue = 0 and red = 0.008). The plots are interpolated from a coarser grid (about 1 cm resolution along the Cartesian axes).

technique is effective in removing the noise from the spatial locations outside the region of interest of the coherence’s peak.

3.2.2. The Results of the Nulling Beamformer. The nulling beamformer is effective in the suppression of the main CMC source as visible in Figure 5. The plots in panel (a) represent the coherence maps of all good subjects (from left—subject 1 to right—subject 4) in a coronal projection, before the suppression (top row) and after the suppression (lower row). The maps are thresholded by a minimum level of coherence, corresponding to a 5% statistical significance, as calculated by the randomization test previously described. It is noteworthy to mention that the newly calculated coherence threshold is smaller than the theoretical one as in (9) by a factor of 5 (e.g., theoretical = 0.01, randomized = 0.002). Figure 5(b) represents the $P^* = (1 - P)$ value map for subject 2, as calculated from the randomization test. The P value map is used to calculate the threshold for the coherence maps (the minimum of coherence among all significant P^* -map voxels constitutes the randomized threshold).

4. Conclusions and Discussion

The application of the proposed analysis pipeline for the localization of CMC sources has evidenced the presence of main CMC activations and putative secondary sources in simulations and real data. Despite the difficulties of

beamformer approaches to correctly localize multiple phase-locked sources, (a phenomenon called source cancellation and demonstrated in formula (29) of VanVeen’s paper, [16]), the application of the here proposed sequential procedure is able to reconstruct the underlying individual sources. In case of a strong primary and a weak secondary source, convincing results have been obtained. In case of similarly strong sources a modified procedure might be required.

The metric of coherence is computationally efficient for the localization of motor sources and can be equally applied to any protocol in which the external signal induces a phase-locked response in the brain. The rationale of applying the described pipeline is the need for a technique that identifies the presence (or absence) of more than one source correlated with EMG. The further step could be the use of these “seeds” for the identification of additional neural aggregates taking part in the motor processes.

Analyses on the single subject level are evidently able to localize the involved oscillatory sources. Although long recording times are needed to improve the SNR and therefore the estimation of the correct beamformer weights, the recording sessions could be divided in training and test sessions, where the first is used to localize the CMC sources and the second to track their time courses. Having a good estimate of the beamformer weights means being able to efficiently extract the activity of the underlying sources and therefore accomplish the ambitious task of a real-time connectivity analysis between EMG and MEG sources.

Based on the significant cortico-muscular coherence found in the present study we cannot unambiguously tell which brain region is the driving force for motor control. Muscular control could be due exerted through (1) both ipsi- and contralateral sources, (2) mainly a contralateral source that drives peripheral muscles directly from contralateral motor cortex and at the same time indirectly via the ipsilateral motor cortex, and (3) another location that drives both contra- and ipsilateral cortices and eventually the peripheral muscles. There is indication that the cerebellum could be involved in this network as well (see lower plots in Figure 5). On this aspect the results are inconclusive and only evidence a clear ipsilateral secondary source. According to this evidence the claims about a functional relation between the sources remain merely speculative. The application of methods like directed coherence or partial directed coherence might be capable to identify the causal flow of information in the network.

The physiological significance of our findings is that during an isometric contraction task multiple cortico-motor sources are recruited that constitute a putative network of motor coordination.

In conclusion, the application of advanced and time-efficient techniques for source suppression/enhancement has made possible the identification of cortico-muscular brain activations that otherwise would not be localizable with standard techniques.

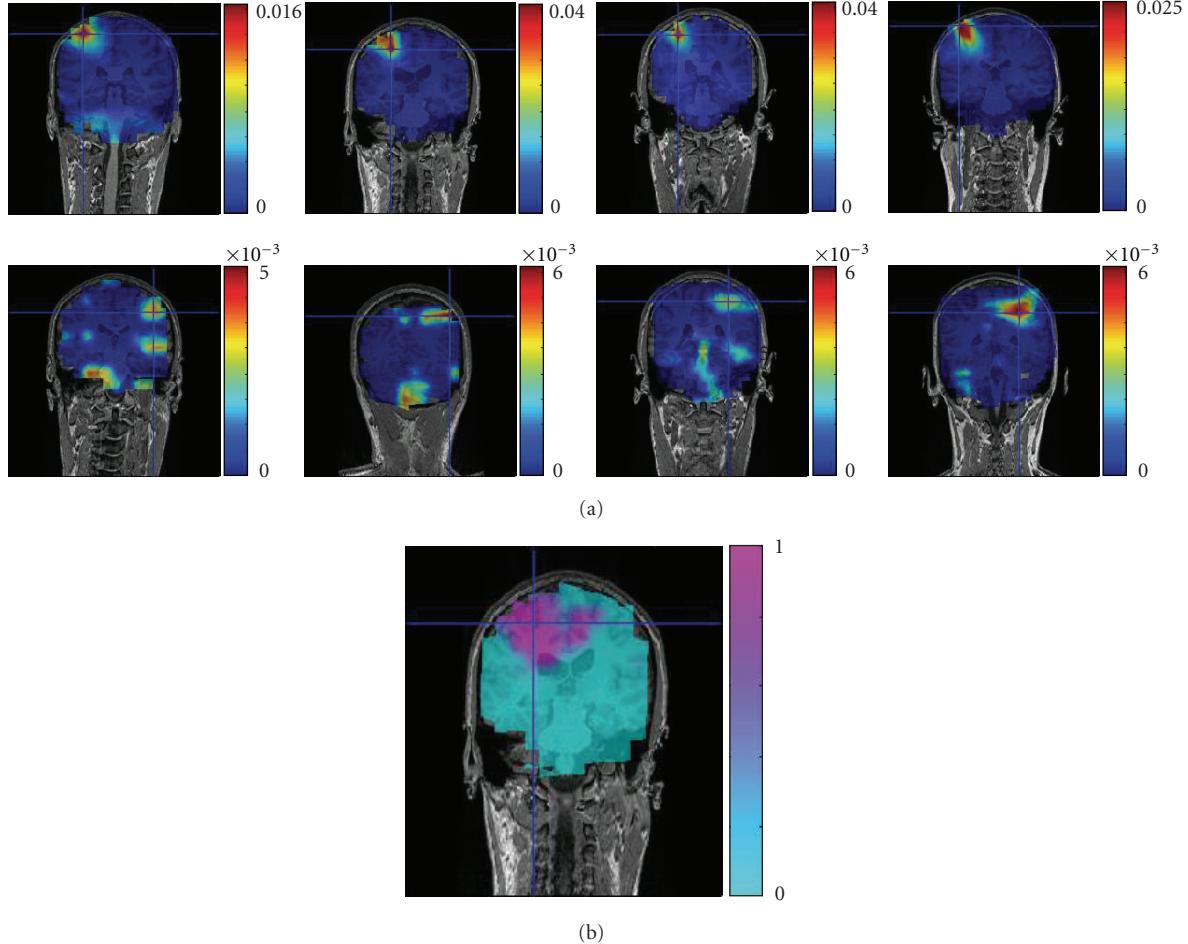


FIGURE 5: (a) The panel shows the results of DICS approach in a coronal slice. The resulting maps are used for the localization of the main source (upper row) and of the secondary source (lower row) in four of the five subjects. (b) The panel shows the map of the P^* values ($P^* = 1 - P$) for subject 2, as a result of a randomization test on the trials of the coherence ($N_{\text{perm}} = 100$). All plots are interpolated versions of a coarser grid (about 1 cm grid resolution along the Cartesian axes).

Acknowledgments

The authors would like to thank Dr. Marco Congedo and Dr. Jan-Mathijs Schoffelen for the useful discussions and Jürgen Dax and Dr. Bianca de Haan for the assistance in the MEG and MRI recordings. Additionally, they would like to thank Werner Reichardt Centre for Integrative Neuroscience in Tübingen and the BrainGain Dutch Consortium for allowing the present paper to be openly accessible.

References

- [1] A. Schnitzler and J. Gross, "Normal and pathological oscillatory communication in the brain," *Nature Reviews Neuroscience*, vol. 6, no. 4, pp. 285–296, 2005.
- [2] M. J. Brookes, C. M. Stevenson, G. R. Barnes et al., "Beamformer reconstruction of correlated sources using a modified source model," *NeuroImage*, vol. 34, no. 4, pp. 1454–1465, 2007.
- [3] H. B. Hui, D. Pantazis, S. L. Bressler, and R. M. Leahy, "Identifying true cortical interactions in MEG using the nulling beamformer," *NeuroImage*, vol. 49, no. 4, pp. 3161–3174, 2010.
- [4] K. Jerbi, J. P. Lachaux, K. N'Diaye et al., "Coherent neural representation of hand speed in humans revealed by MEG imaging," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 18, pp. 7676–7681, 2007.
- [5] K. Sekihara and S. S. Nagarajan, *Adaptive Spatial Filters for Electromagnetic Brain Imaging*, Springer, Berlin, Germany, 2008.
- [6] S. S. Dalal, K. Sekihara, and S. S. Nagarajan, "Modified beamformers for coherent source region suppression," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1357–1363, 2006.
- [7] J. M. Schoffelen, R. Oostenveld, and P. Fries, "Imaging the human motor system's beta-band synchronization during isometric contraction," *NeuroImage*, vol. 41, no. 2, pp. 437–447, 2008.
- [8] I. T. Chen, Y. S. Chen, and L. F. Chen, "Beamformer-based spatiotemporal imaging of correlated brain activities," in *9th IEEE International Conference on Bioinformatics and BioEngineering (BIBE '09)*, pp. 431–434, June 2009.

- [9] S. S. Dalal, A. G. Guggisberg, E. Edwards et al., "Spatial localization of cortical time-frequency dynamics," in *29th Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBC '07)*, pp. 4941–4944, August 2007.
- [10] M. J. Brookes, C. M. Stevenson, A. Hadgipapas et al., "Beamformer imaging and interference rejection," in *Biomagnetism. Interdisciplinary Research and Exploration*, R. Kakigi, K. Yokosawa, and S. Kuriki, Eds., pp. 62–64, Hokkaido University Press, Sapporo, Japan, 2008.
- [11] A. Hillebrand and G. R. Barnes, "Beamformer Analysis of MEG Data," *International Review of Neurobiology*, vol. 68, pp. 149–171, 2005.
- [12] B. A. Conway, D. M. Halliday, S. F. Farmer et al., "Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man," *Journal of Physiology*, vol. 489, no. 3, pp. 917–924, 1995.
- [13] T. Mima and M. Hallett, "Corticolumbar coherence: a review," *Journal of Clinical Neurophysiology*, vol. 16, no. 6, pp. 501–511, 1999.
- [14] J. M. Kilner, S. N. Baker, S. Salenius, R. Hari, and R. N. Lemon, "Human cortical muscle coherence is directly related to specific motor parameters," *Journal of Neuroscience*, vol. 20, no. 23, pp. 8838–8845, 2000.
- [15] D. M. Halliday, B. A. Conway, S. F. Farmer, and J. R. Rosenberg, "Using electroencephalography to study functional coupling between cortical activity and electromyograms during voluntary contractions in humans," *Neuroscience Letters*, vol. 241, no. 1, pp. 5–8, 1998.
- [16] B. D. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 867–880, 1997.
- [17] K. Sekihara, S. S. Nagarajan, D. Poeppel, and A. Marantz, "Performance of an MEG adaptive-beamformer technique in the presence of correlated neural activities: effects on signal intensity and time-course estimates," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 12, pp. 1534–1546, 2002.
- [18] J. Gross, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, and R. Salmelin, "Dynamic imaging of coherent sources: studying neural interactions in the human brain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 2, pp. 694–699, 2001.
- [19] M. Congedo, "Subspace projection filters for real-time brain electromagnetic imaging," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 8, pp. 1624–1634, 2006.
- [20] A. Rodríguez-Rivera, B. V. Baryshnikov, B. D. Van Veen, and R. T. Wakai, "MEG and EEG source localization in beamspace," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 430–441, 2006.
- [21] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP magazine*, vol. 5, no. 2, pp. 4–24, 1988.
- [22] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "Field-Trip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 156869, 9 pages, 2011.
- [23] K. Sekihara, M. Sahani, and S. S. Nagarajan, "Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction," *NeuroImage*, vol. 25, no. 4, pp. 1056–1067, 2005.
- [24] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*, Elsevier/Academic Press, Burlington, Mass, USA, 2005.
- [25] G. Nolte, "The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors," *Physics in Medicine and Biology*, vol. 48, no. 22, pp. 3637–3652, 2003.
- [26] A. Gramfort, T. Papadopoulou, E. Olivi, and M. Clerc, "Forward field computation with OpenMEEG," *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 923703, 13 pages, 2011.
- [27] "Tutorial on artifact detection," http://fieldtrip.fcdonders.nl/tutorial/automatic.artifact_rejection.
- [28] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *NeuroImage*, vol. 34, no. 4, pp. 1443–1449, 2007.
- [29] O. P. Neto and E. A. Christou, "Rectification of the EMG signal impairs the identification of oscillatory input to the muscle," *Journal of Neurophysiology*, vol. 103, no. 2, pp. 1093–1103, 2010.
- [30] P. P. Mitra and B. Pesaran, "Analysis of dynamic brain imaging data," *Biophysical Journal*, vol. 76, no. 2, pp. 691–708, 1999.
- [31] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG- and MEG-data," *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, 2007.
- [32] CTF head coordinates conventions, "CTF systems," Head Localization, manual number 900-0017-AW02 release 3.2, 93–95, 1993.

Research Article

Computational Models of Neuron-Astrocyte Interactions Lead to Improved Efficacy in the Performance of Neural Networks

Alberto Alvarellos-González, Alejandro Pazos, and Ana B. Porto-Pazos

Departamento de Tecnologías de la Información y las Comunicaciones, Facultad de Informática, Universidad da Coruña, Campus de Elviña, 15071 A Coruña, Spain

Correspondence should be addressed to Ana B. Porto-Pazos, ana.portop@udc.es

Received 14 October 2011; Revised 12 January 2012; Accepted 25 February 2012

Academic Editor: Luca Faes

Copyright © 2012 Alberto Alvarellos-González et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The importance of astrocytes, one part of the glial system, for information processing in the brain has recently been demonstrated. Regarding information processing in multilayer connectionist systems, it has been shown that systems which include artificial neurons and astrocytes (Artificial Neuron-Glia Networks) have well-known advantages over identical systems including only artificial neurons. Since the actual impact of astrocytes in neural network function is unknown, we have investigated, using computational models, different astrocyte-neuron interactions for information processing; different neuron-glia algorithms have been implemented for training and validation of multilayer Artificial Neuron-Glia Networks oriented toward classification problem resolution. The results of the tests performed suggest that all the algorithms modelling astrocyte-induced synaptic potentiation improved artificial neural network performance, but their efficacy depended on the complexity of the problem.

1. Introduction

The behavior of the nervous system (NS) remains a mystery in many respects. The details of how the brain performs certain information processing tasks, such as classification, pattern recognition, and concept abstraction, are still unknown. Although it has long been thought that neurons were the only cells involved in complex cognitive processes, this thinking has changed. Recent discoveries show the importance of particular glial cells, called astrocytes, for information processing in the brain [1–8]. Abundant evidence suggests the existence of bidirectional communication between astrocytes and neurons and an important active role for the astrocytes in the physiology of the NS [1, 3–5]. This evidence has led to the proposal of a new concept in synaptic physiology, the tripartite synapse, which consists of three functional elements: the presynaptic and postsynaptic elements, and the surrounding astrocytes [2]. The communication between these three elements has highly complex characteristics, which seem to reflect more reliably the complexity of the information processing between elements of the NS. In order to understand the motives

of this reciprocal signalling, we must know the differences and similarities that exist between their properties. Only a decade ago, it would have been absurd to suggest that these two cell types have very similar functions; now we realise that the similarities are striking from the perspective of chemical signalling. Both cell types receive chemical inputs that have an impact on their ionotropic and metabotropic receptors. Following this integration, both cell types send signals to their neighbours through the release of chemical transmitters. Both neuron-to-neuron signalling and neuron-to-astrocyte signalling show plastic properties that depend on the particular activity [1]. The main difference between astrocytes and neurons is that many neurons extend their axons over large distances and conduct action potentials of short duration at high speed, whereas astrocytes do not exhibit any electric excitability but conduct calcium spikes of long duration (tens of seconds) over short distances and at low speed. The fast signalling, and input/output functions in the central NS that require speed seem to belong to the neural domain. But what happens with slower events, such as induction of memories and other abstract processes, such as thought processes? Does signalling between astrocytes

contribute to their control? As long as there are no answers to these questions, research must continue; the present work offers new ways to advance through the use of Artificial Intelligence (AI) techniques.

This work tries to add some new knowledge about the interaction of neurons and astrocytes regarding information processing, in both the brain and in computer AI systems. Hence, this is a multidisciplinary study. It tries to benefit both Neuroscience, by helping to understand the neuron-glia interaction, and AI, creating new computational methods for processing information. Including artificial elements that attempt to imitate astrocytes' behavior in Artificial Neural Networks (ANNs) has proven to present advantages in classification problems [9]. This inclusion gave rise to the so-called *Artificial Neuron-Glia Networks* (ANGNs) [9, 10]. In our previous work we have investigated the consequences of including artificial astrocytes, which mimic biologically defined properties involved in astrocyte-neuron communication, on artificial neural network performance. Using connectionist systems and evolutionary algorithms, we have compared the performance of ANN and ANGN in solving classification problems. We have shown that the degree of success of ANGN was superior to that of ANN. Analysis of the performance of ANN with different numbers of neurons or different architectures indicated that the effects of ANGN cannot be accounted for by an increased number of network elements but rather are specifically due to astrocytes. Furthermore, the relative efficacy of ANGN versus ANN increased as the complexity of the network increased [9]. It is important to note that our AI computational model does not account for the primary intrinsic physiological property of astrocytes, intercellular calcium waves in the astroglial network, such as in the work of Ikuta et al. [11, 12]. Instead of building an astroglial network with intercellular waves parallel to the neuronal network, and then analyzing their conjoint operation, we added single astrocytes to single neurons, allowing the astrocyte to increase the strength of the connections of the neuron with the next neuronal layer (see Section 2). At present, we are also modelling other types of astrocytic influence.

The neuron-astrocyte interaction in the ANGN implemented in our previous work constituted what we have called a neuron-glia algorithm. That first algorithm, which we have named *Attenuated effect of astrocyte* (see Section 3), tried to imitate a behavior observed between neurons and astrocytes in the hippocampus [6]. However, several mechanisms and physiological consequences of astrocyte-neuron communication occur in the brain. Under what conditions one specific modulatory effect takes place in a particular neural network remains unknown [8]. Therefore, in the present work we have researched whether other forms of interaction may help us understand what happens in the brain, and which may prove to be more or equally efficient in information processing in computers. We have modeled new and different neuron-astrocyte interactions, analyzing the results of Neuroscience experiments carried out in biological nervous systems [4–8]. This has led to the implementation of different neuron-glia algorithms for training and validation of feed-forward multilayer ANGN used in resolution of

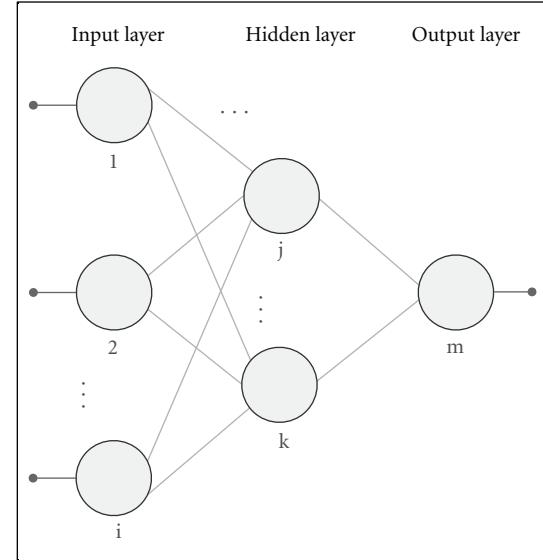


FIGURE 1: Artificial Neural Network structure.

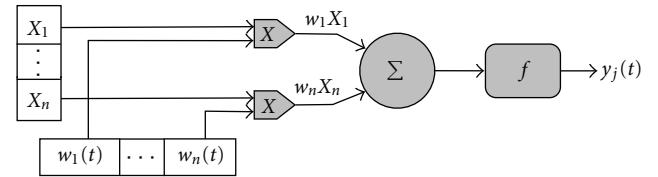


FIGURE 2: Structure of an artificial neuron.

classification problems. For simplicity, our work focused on modelling astrocyte-induced synaptic potentiation, as in our previous study [9]. The results obtained using these new algorithms allowed comparisons between them and the observation of neuron-glia interactions that, so far, obtain the best results for each problem.

This paper is organised into the following sections. Section 2 introduces the ANGN and explains its overall behavior, and its differences with multilayer networks without artificial astrocytes. Section 3 details the implementation of neuron-glia algorithms created for the study of neuron-astrocyte interaction. Section 4 explains the simulations performed applying the created algorithms to two problems, *simulation of a multiplexor device* and *iris flower classification*, and shows the results obtained from these simulations. Finally, Section 5 summarizes the discussion and conclusions of this study and explains the work that is being developed as a continuation of this research.

2. Artificial Neuron-Glia Networks

ANNs are interconnected neuron models that simulate the behavior of biological neural networks [13, 14] (see Figure 1). The neuron is the basic information-processing unit in these networks (see Figure 2).

The connectionist branch of AI carries out the study of ANN. Researchers in this area have designed and built

different types of ANN; these systems are different in their topology, dynamics, and behavior of their constituent elements separately and together (the system as a whole). Many advances have been made in those aspects, but there are many limitations in the areas of processing speed and computational complexity of the connectionist systems.

The connection between two neurons is a directional one. Hence, one neuron a is the source of the connection and the other b is the destination neuron. A value is associated with the connection. This value is known as *weight* and it determines the influence of the connection in the activation of the destination neuron.

A neuron j , is characterized by n inputs, with signals x_1 to x_n and weights w_1 to w_n associated with the inputs.

The signals may come from other neurons or may be the input signals of the network. The output of the neuron j is given by the application of the transfer function $f : \mathbb{R} \rightarrow \mathbb{R}$ [15] to the sum of the inputs adjusted by its associated weight:

$$y_j(t) = f\left(\sum_{i=1}^n w_i x_i\right). \quad (1)$$

An ANGN extends the ANN architecture by including a new kind of processing element, the artificial astrocyte [9, 10, 16–18] (see Figure 3).

An artificial astrocyte is associated with a neuron and controls its activity (see Figure 4). The astrocyte modifies the weight of the connections it is associated with (input connections, output connections, or both), depending on the activity of the neuron. A range of values is also associated with a weight, known as a weight limit (wM). The astrocyte controls the activity of the j neuron by using a counter. That counter records the times that the neuron fires.

Due to the lack of knowledge regarding the specific characteristics of the modifications that astrocytes make in neuronal connections, we implemented different *neuron-glia algorithms* (see Section 3) to simulate the behavior which astrocytes of the brain are presumed to have, considering the observations made on the nervous systems of living organisms [1–4]. Glutamate released in the extracellular space by an astrocyte or a presynaptic neuron can affect another astrocyte, another presynaptic neuron, or a postsynaptic neuron. If the glutamate that reaches a postsynaptic neuron proceeds directly from a presynaptic neuron, the action potential (AP) takes place more rapidly and ends more or less quickly. If there has also been a release of glutamate by an astrocyte that was activated by the glutamate of a presynaptic neuron, more AP will take place [1]. The activation of astrocytes is a slow process, if we compare it with neural activity [4]. The same conclusion can be drawn from their effect on the synapse between two neurons, whose neurotransmitters activated the astrocyte, and which is one thousand times slower than the propagation of the impulse by the neurons. This slowness has led to the presentation to the ANGN of each training pattern during more than one cycle or iteration. If it imitates this slowness, the ANGN will need k cycles or iterations to process each input pattern (see Figure 4).

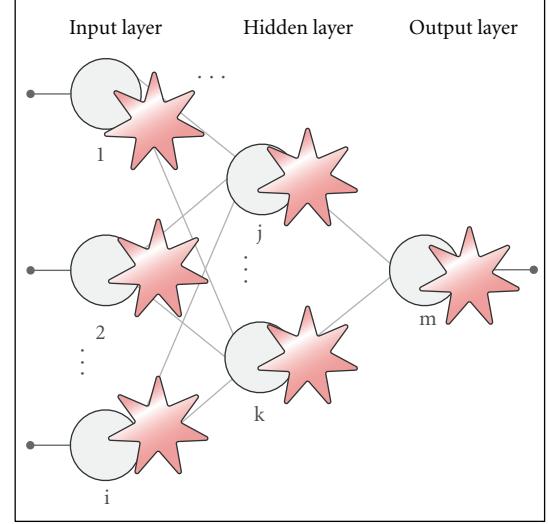


FIGURE 3: Artificial Neuron-glia network structure.

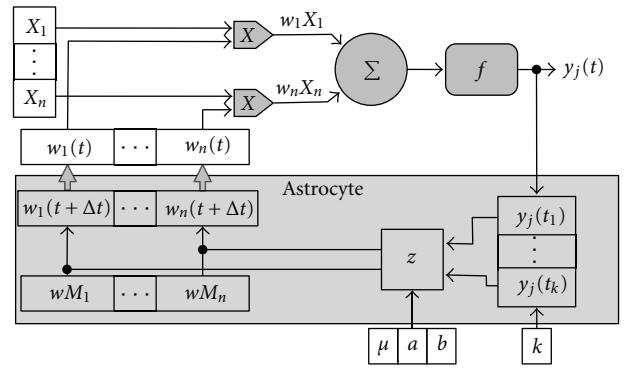


FIGURE 4: Astrocyte representation.

We must also consider that the contribution of the astrocytes to the weights of the ANGN connections takes place according to a time factor, given the fact that they act slowly and their responses are non-linear. It would be interesting to know how astrocytes affect the connectionist system, considering their influence on the synapses according to the activities of the neurons over the course of time. The more intense the activity of the neurons, the larger the influence of the astrocyte on a connection.

The behavior of an astrocyte is determinate by the parameters $k \in \mathbb{N} \setminus \{0\}$, $\mu \in [1, k]$ and $a, b \in [0, 1]$. Each instance or input pattern that is used for training, validating or testing the artificial net is processed k times (iterations). The astrocyte registers the activity of the neuron during the k iterations, applying a function $u : \mathbb{R} \rightarrow \mathbb{Z}$ over the output of the neuron $y_j(t)$, where u indicates if the neuron has fired $u(x) = 1$ or not $u(x) = -1$:

$$u(x) = \begin{cases} -1, & x \leq 0, \\ 1, & x > 0. \end{cases} \quad (2)$$

Hence the astrocyte has a register of the neuron's activity with a temporal window of k instants of time (an iteration lasts

one instant of time). Observing this activity, the astrocyte will modify the weight of its associated neuronal connections when the counter of the activity of the neurons reaches the value μ . Figure 4 shows how the input neuronal connections are modified. An astrocyte may also modify output neuronal connections or both

$$w_i(t + \Delta t) = w_i(t) + \Delta w_i(t), \quad (3)$$

where

$$\Delta w_i(t) = |w_i(t)| z(t), \quad (4)$$

and $z : \mathbb{N} \setminus \{0\} \rightarrow \mathbb{R}$ is a function defined as

$$z(t) = \begin{cases} a, & r_j(t) = \mu, \\ -b, & r_j(t) = -\mu, \end{cases} \quad (5)$$

with $r_j : \mathbb{N} \setminus \{0\} \rightarrow [-\mu, \mu]$ being the function that returns the number of times a neuron has fired. If the neuron was active μ times, the weights of the connections will be increased by a percentage a , while they will be decreased by a percentage b if the neuron remained inactive during μ iterations.

3. Neuron-Glia Algorithms

The six algorithms implemented were different in two aspects: the specific implementation they make of the r_j function, and whether or not they respect the weight limit when the neuronal connection is being modified. The different implementations of the r_j function of each algorithm are explained in its corresponding subsection. Different approaches regarding the modification of the connections weight are also explained.

3.1. Consecutive Activations, Weight Limited. The astrocyte respects the weight limit of the connections:

$$w_i(t + \Delta t) = \min \{w_i(t) + \Delta w_{ji}(t), wM_i\}. \quad (6)$$

This algorithm contemplates only consecutive neuron activations; if the neuron reaches the activity or inactivity level that makes the astrocyte act, the activity counter is restarted. The neuronal activity level, following these restrictions, is given by the following function:

$$r_j(t) = \begin{cases} u(y_j(t)) + r_j(t-1), & t > 0, u(y_j(t)) = u(y_j(t-1)), \\ & r_j(t-1) \in (-\mu, \mu), \\ u(y_j(t)), & \text{in other case.} \end{cases} \quad (7)$$

3.2. Consecutive Activations, Weight Unlimited. The behavior of this algorithm is the same as the previous one, except that in this case the astrocyte will not respect the limit weight of the connections; hence they can reach any value:

$$w_i(t + \Delta t) = w_i(t) + \Delta w_i(t). \quad (8)$$

3.3. Nonconsecutive Activations, Weight Limited. The astrocyte respects the weight limit of the connections

$$w_i(t + \Delta t) = \min \{w_i(t) + \Delta w_{ji}(t), wM_i\}. \quad (9)$$

In this algorithm the neuron activations need not be consecutive. If the neuron reaches the activity or inactivity level that makes the astrocyte act, the activity counter is restarted. The neuron activity level, following these restrictions, is given by the following function:

$$r_j(t) = \begin{cases} u(y_j(t)) + r_j(t-1), & t > 0, r_j(t-1) \in (-\mu, \mu), \\ u(y_j(t)), & \text{in other case,} \end{cases} \quad (10)$$

Having the activity of the neuron not required to be consecutive gives rise to this result: if an astrocyte increments the weight of a connection of a neuron, it indicates that the neuron fired μ iterations more than it remained inactive. If an astrocyte decrements the weight of a connection to a neuron, it indicates that the neuron fired μ iterations less than it remained inactive.

3.4. Nonconsecutive Activations, Weight Unlimited. The behavior of this algorithm is the same as the previous one, except that in this case the astrocyte will not respect the limit weight of the connections; hence they can reach any value:

$$w_i(t + \Delta t) = w_i(t) + \Delta w_i(t). \quad (11)$$

3.5. Attenuated Effect of Astrocyte. In this algorithm, the astrocyte will not respect the limit weight of the connections:

$$w_i(t + \Delta t) = w_i(t) + \Delta w_i(t), \quad (12)$$

and the activity of the neuron need not be consecutive

$$r_j(t) = \begin{cases} u(y_j(t)) + r_j(t-1), & t > 0, r_j(t-1) \in (-\mu, \mu), \\ r_j(t-1), & t > 0, r_j(t-1) \in \{-\mu, \mu\}, \\ u(y_j(t)), & \text{in other case.} \end{cases} \quad (13)$$

The major difference with the previous algorithms stems from the management of the activity counter of the neuron: when the neuron reaches the activity level $\{-\mu, \mu\}$ that makes the astrocyte modify its neuronal connections, the activity counter is not set to zero (it retains the value). This behavior has a noticeable consequence in the modification of the connections weight: when the point at which an astrocyte modifies the weight of the connections is reached in a given iteration and the neuron fires again in the next iteration, the astrocyte will increase the connections weight of the neuron again. The behavior when the neuron remains inactive is similar, with the outcome being the weight is decreased.

In the previous algorithms, having the activation counter be set to zero, the counter needed to reach the value $\{-\mu, \mu\}$ again for the astrocyte to act (thus a minimum of μ iterations

of neuronal activity/inactivity are required). This behavior implies an extra reinforcement on those neurons that fire the most, it also makes the astrocytic effect last longer, and disappear only gradually over time.

3.6. Global Processing Effect. In the previous algorithms, each instance was processed a certain number of iterations to try to simulate the delay in glial system functioning (an order of magnitude of seconds) with respect to neuronal functioning (an order of magnitude of milliseconds) (see Figure 5) [19, 20].

This algorithm named *Global processing effect* was created under the assumption that the way the brain works is different than the way it was being simulated. Instead of processing each instance during k iterations, this algorithm considers the instances as a whole, and the net processes the whole set of instances during k iterations (see Figure 6). For example, this algorithm considers the following: when visual information (i.e., a scene) is being processed by the brain, it obtains a sequence of images and this sequence is processed as a whole. Common characteristics are extracted from these images to give the scene meaning. The previous algorithms would consider each image independently and would try to extract characteristics from each one separately.

The way this new algorithm modifies the weights of the connections is different from the previous algorithms. with where n is the number of instances (for training/validating/testing), this algorithm evaluates the n instances during k iterations. The weights are modified using

$$w_i(t + \Delta t) = w_i(t) \Delta w_i(t), \quad (14)$$

where

$$\Delta w_i(t) = \frac{n + \text{sgn}(w_{ji}(t))s}{n}, \quad (15)$$

With s being the number of times the neuron was active during k iterations. With this behavior, the modification of connections weight of the neuron is directly proportional to the neuron's activity.

4. Results

To evaluate the functioning of the neuron-glia algorithms, ANGN using our different algorithms was compared with ANN (without artificial astrocytes) trained only by using Genetic Algorithms (GAs). The ANGN training method is a hybrid one. It is composed of two learning phases: an unsupervised learning phase (where a selected neuron-glia algorithm is used) and a supervised learning phase (where GAs are applied using the MSE calculated in the unsupervised phase) [9, 10, 16]. The networks to be compared (seven networks: ANGN with 6 different unsupervised algorithms and the ANN—see Table 1) were trained to solve two classification problems of increasing complexity. The problems were taken from the UCI Machine Learning Repository [21]: the *simulation of a multiplexed device* problem and the *Iris Flower* problem.

TABLE 1: Methods name summary.

Method	Name
1	Consecutive, weight limited
2	Consecutive, weight unlimited
3	Not consecutive, weight limited
4	Not consecutive, weight unlimited
5	Attenuated effect of astrocyte
6	Global processing effect
7	ANN

The network architecture chosen for each problem was selected based on good results achieved in previous work by Rabuñal et al. [22, 23]. Anyway, what matters is not to have the best architecture. The important issue is to have the same architecture in all the networks to be compared, in order to test effects caused only by the inclusion of artificial astrocytes. Anyway, we had shown in our previous work [9] that by increasing the number of neurons and layers, the effect of artificial astrocytes becomes even more beneficial.

Regarding the parameters of neuron-glia algorithms in ANGN, four different combinations of iterations k (4, 6, or 8) and activations μ (2 or 3) were considered for each algorithm, in particular (4-2, 6-2, 6-3 y 8-3); after doing preliminary tests [16], the combinations that obtained the best results were selected. The same parameters (k , μ , a , and b) were used for all the artificial astrocytes in each simulation. Each simulation also used a 25% weight increment (a) and a 50% weight decrement (b). This decision was based on the good results obtained in previous work [16]. Biological knowledge supports this choice, because a lower increment than a decrement, when there is no constant activity, reinforces the connections of those neurons that show a constant activity [24].

Regarding the GA parameters, they were also chosen in contemplation of previous work from Rabuñal et al. [22, 23]. Those parameters were not intended to obtain the best results in every problem but to use the same parameters in the connectionist systems to be compared. A population of 100 individuals was used in the GA. The individual selection algorithm chosen was “Montecarlo” and the substitution method was the Darwinian substitution. Regarding individual breeding, a single crossover point was used. The crossover rate was set to 90% and the mutation rate to 10%.

The simulations were performed by means of a tool we have implemented with Borland DELPHI and Visual C++ languages. The tests were run in an AMD Athlon PC, with 1 GB of RAM and Windows XP OS.

4.1. Simulation of a Multiplexor Device (MUX). We decided to carry out the first tests over a well-known multilayer architecture, which solves a simple classification problem: the simulation of an electronic device called MULTipleXor (MUX) with four inputs and one output (see Figure 7). For the resolution of this problem a three-layer network was used: six neurons in the input layer, four in the hidden layer

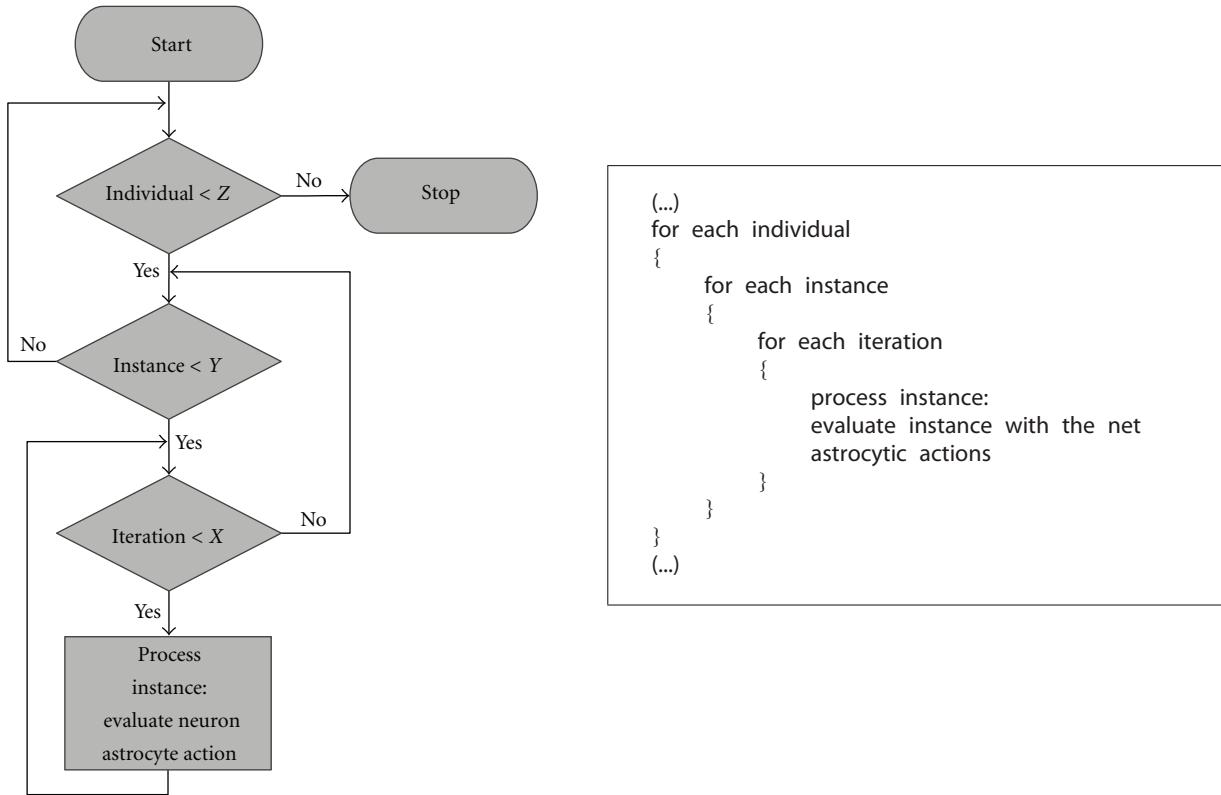


FIGURE 5: Flow chart and pseudocode of the five Neuron-Glia algorithms.

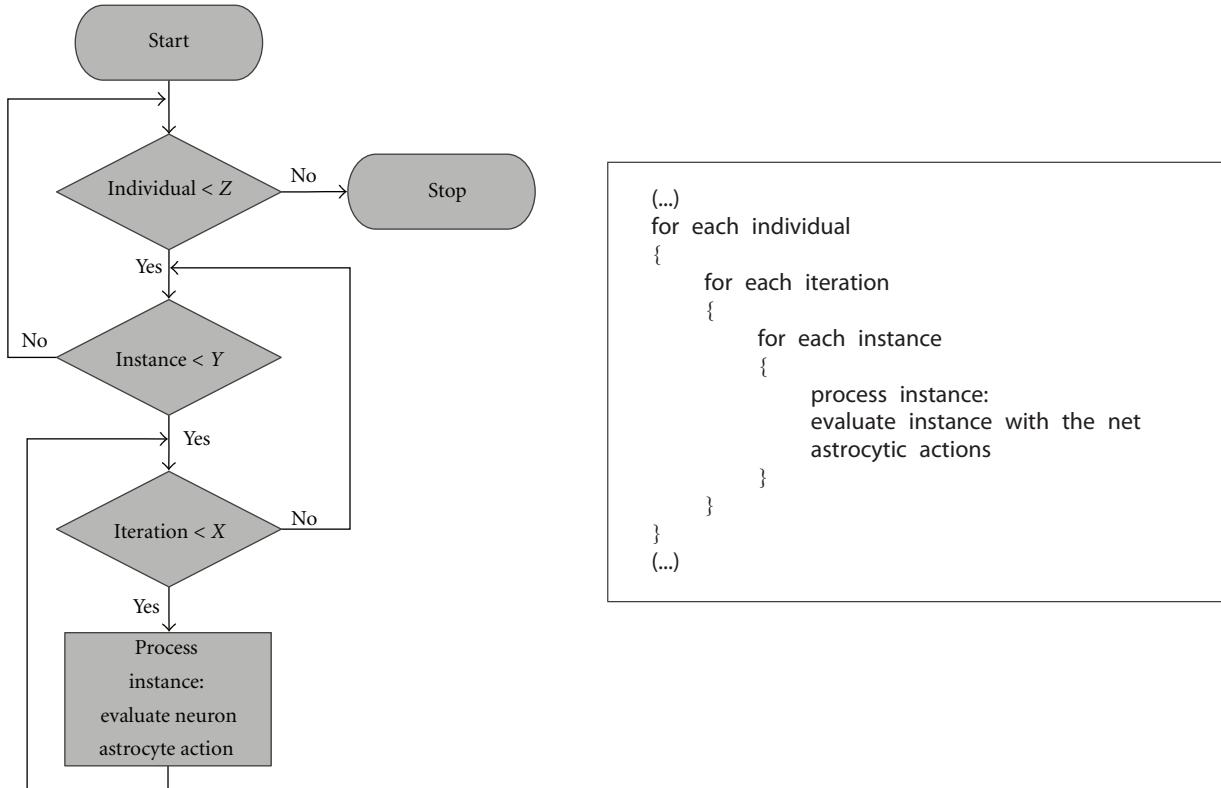


FIGURE 6: Global processing effect flowchart and pseudocode.

TABLE 2: MUX mean results.

Method	Generation	Training error (ECM)	Training standard deviation (%)	Validation accuracy (%)	Validation standard deviation	Time
1	222	0,132	0,056	86,25	8,75	0:00:21
2	282,1	0,091	0,027	81,25	10,08	0:00:38
3	345,2	0,125	0,059	86,25	8,75	0:00:29
4	355,6	0,083	0,021	81,25	10,08	0:00:45
5	681,3	0,108	0,060	86,25	10,38	0:01:08
6	563,6	0,096	0,057	76,25	3,75	0:00:33
7	521,8	0,101	0,051	62,5	9,68	0:00:07

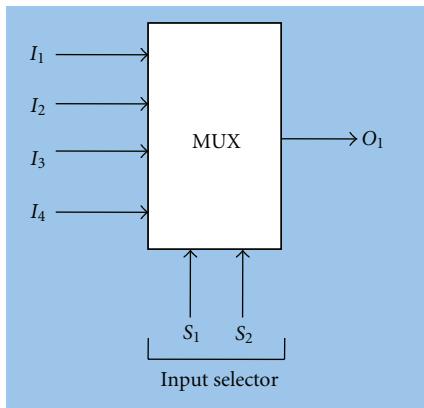


FIGURE 7: MUX device.

and one neuron in the output layer. The activation function used for all the neurons in the network was a threshold function with a threshold value $\theta = 0.5$. The output values of this problem are boolean. The limit weight for all the neurons was set to one. The 64 instances available for this problem were divided into two sets: 58 instances for training, and six instances where different classes were selected for the purpose of checking the generalization capacity. All simulations were executed over 4000 generations.

All implemented neuron-glia algorithms were tested with the MUX problem (see Table 1).

Each algorithm was executed using ten different populations of connection weights. For each population and algorithm the best values (those with higher validation accuracy) were chosen from among the aforementioned four μ and k combinations used. Table 2 shows the mean values for each algorithm. These results show that in all cases but two, ANGN achieved a lower training error, and in all cases a higher validation accuracy was achieved by ANGN with respect to ANN results.

The number of times each algorithm achieved the best results for each measurement was analyzed (fewer generations, lower training error, higher validation accuracy, or less time). Table 3 shows that ANGN achieved the best results more frequently, considering the validation accuracy, despite taking more time to achieve it.

4.2. Iris Flowers Classification (Iris). We wanted to prove our algorithms with a problem related to a much more complex

TABLE 3: MUX results summary.

Method	Generation	Training error	Validation accuracy	Time	Total
1	3	1	3	1	8
2	2	3	3	1	9
3	3	1	3	1	8
4	2	3	3	1	9
5	1	3	4	0	8
6	0	4	0	3	7
7	3	3	0	6	12

domain than MUX, a problem to test the algorithms where the ANGN is dealing with multiple classification tasks. In contrast to the MUX problem, the IRIS flower problem uses continuous input values, different activation functions in artificial neurons of different layers, and twice as many training patterns. It consists in identifying a plant's species: Iris setosa, Iris versicolor, or Iris virginica. This case has 150 examples with four continuous inputs which stand for four features about the flower's shape. The four input values represent measurements in millimetres of petal width, petal length, sepal width, and sepal length. The learning patterns have been found to have four inputs and three outputs. The three outputs are Boolean ones, representing each Iris species. By doing it in this manner (three boolean outputs instead of a multiple one), additional information can be provided about whether the system's outputs are reliable or not. That is, due to the outputs' intrinsic features, only one of them must possess the true value, standing for the type of flower it has classified, while the rest have a false value. Therefore, if two or more outputs have true values, or if all of them are false, we may conclude that the value classified by the system is an error and the system cannot classify that case. The values corresponding to the four input variables have been normalized in the interval (0–1) so that they are dealt with by the ANGN.

For the resolution of this problem a three-layer network was used: four neurons in the input layer, five in the hidden layer, and three neurons in the output layer (one neuron per iris class). The activation function used for all the neurons in the network was a hyperbolic tangent, except for the output layer neurons, where a threshold function with a threshold value $\theta = 0.5$ was used. The limit weight for all the neurons

TABLE 4: Iris methods summary.

Method	Name
1	Not consecutive, weight unlimited
2	Attenuated Effect of Astrocyte
3	Attenuated Effect of Astrocyte 2
4	ANN

was set to 1. The 150 instances available for this problem were divided into two sets: 2/3 for training and 1/3 for validation; the training set is composed of 1/3 of instances of each class. All simulations were executed over 2000 generations.

Although, as mentioned previously, what matters is having the same architecture, we have established this architecture and these parameters, which were obtained by our research group in previous work [22, 23]. By using ANN with these features, and training exclusively by means of GA, Rabuñal et al. [22, 23] reached an adjustment better than the previous best example of work for solving the IRIS flower with ANN, in which Martínez and Goddard [25] used BP for the training and a hidden neuron more than Rabuñal et al. These good results demonstrated the efficacy of GA for simplifying and solving this problem. We compared our new ANGN with ANN trained exclusively by means of GA.

The algorithm named *Attenuated Effect of Astrocyte* was chosen because it achieved the best results in the MUX problem (good mean validation accuracy, and the best results achieved more often). With the aim of adding some variability to the results, two configurations of this algorithm were used: one with $k \in [4, 6, 8]$ and $\mu \in [2, 3]$, and another with $k \in [3, 5, 7]$ and $\mu \in [2, 3]$. The latter is called *Attenuated Effect of Astrocyte 2* in Table 4.

Another algorithm was chosen from the remaining ones to test its behavior against a more complex problem. The less restrictive algorithm was chosen (in terms of both activations of neurons and restrictions on modifications of weights of connections): “Not consecutive activations, weight unlimited”.

As with the MUX problem, each algorithm was executed using ten different populations of weights. For each population and algorithm the best values (those with higher validation accuracy) were chosen among the four μ and k combinations used. Table 5 shows the mean values for each algorithm. These results show that in all cases but two, ANGN achieved a lower training error, and in all cases a higher validation accuracy was achieved by ANGN with respect to results with ANN.

Regarding the number of times each algorithm obtained the best results, Table 6 shows that ANGN achieved the best results more times, considering the validation accuracy, despite taking more time to achieve it.

5. Discussion and Conclusions

All the implemented algorithms have tried to emulate, in the ANGN, the potentiation of synaptic connections that take place in the brain caused by astrocytes, due to high

synaptic activity. The first five algorithms emulate astrocytic behavior in a similar way, just changing the restrictions on the weights changes and the consecutive or not consecutive nature of synaptic activity. Unlike in the first five algorithms, the sixth algorithm operates in a fairly different manner: it considers global information processing rather than an individual instance.

ANGN implemented using these six algorithms (thus including artificial astrocytes that simulate the potentiation of the connections and penalize the lack of activity) improved the ANN that did not include artificial astrocytes. It is worth noticing the difference in efficacy between the first five algorithms and the sixth, which achieved the worst results.

It was also observed that the *Not consecutive, weight unlimited* algorithm did not achieve the best results in the simpler problem (MUX) but did achieve the best results in the more complex problem (Iris) with a more complex net architecture. This suggests that the net (and problem) complexity influences the behavior of an algorithm, rendering an algorithm more appropriate for one problem or another depending on the problem’s complexity. This allows the conclusion that a specific behavior of astrocytes is better suited to some problems rather than to others, depending on the problem’s complexity and characteristics. This behavior agrees with the biological behavior of astrocytes in the brain. Sometimes they have more influence than at others. Moreover, it has been observed that the number of astrocytes is higher in more complex brain areas, and they have more influence in them. The highest ratio of glia-to-neurons is found at the top of the phylogenetic tree, in the human brain; this leaves us with the question as to whether astrocytes are key regulatory elements of higher cortical functions [3].

In any case, to prove the hypothesis obtained, more tests are being performed with the Iris problem and with other problems (starting from the preliminary test and the algorithms developed in this work). Therefore, the development of models of astrocyte-neuron interaction that incorporate the richness of biological interactions, for example, astrocyte-induced synaptic depression, or depression and potentiation altogether, is being undertaken to test whether they provide similar, or even better, results in neural network performances.

This work attempts to assist both AI and Neuroscience, the former by creating a new kind of information processing element for neural networks, the latter by contributing ideas that hope to, somehow, guide investigations in Neuroscience toward understanding the behavior of astrocytes in the nervous systems of living organisms. All the algorithms presented could attempt to be translated in a laboratory for Neuroscience. They may mimic physiological dynamics that occur in the brain.

Understanding the importance of astrocytes in brain function, and learning how communication occurs between neuronal networks and astrocyte networks at a microscopic level, is crucial for the understanding of the synergic behavior of cerebral regions. It is known that astrocytes are not present in the same proportion in all brain regions, which may influence the existence of different modes of interaction between one and other areas. The behavior of astrocytes in

TABLE 5: Iris mean results.

Method	Generation	Training error (ECM)	Training standard deviation	Validation accuracy (%)	Validation standard deviation	Time
1	693,9	0,065	0,023	78,2	5,76	0:03:17
2	486,8	0,151	0,079	72,4	5,64	0:02:44
3	868,6	0,155	0,098	70,2	8,17	0:04:30
4	166,6	0,371	0,051	56	4,29	0:00:09

TABLE 6: Iris results summary.

Method	Generation	Training error	Validation accuracy	Time	Total
1	0	8	7	0	15
2	0	0	2	0	2
3	1	2	2	1	6
4	9	0	0	9	18

some regions is just beginning to be analyzed, in which its effects on information processing have never been studied, such as cerebral cortex. Computational models to study brain circuitry connectivity at a microscopic level will allow an understanding of what happens at a macroscopic level. Moreover, given the efficacy of ANGN in processing information, they could provide a double benefit to this area of study. Since they can be used to classify and recognize patterns, ANGN will be tested in the near future as a data analysis tool for helping to detect any characteristic or neurological disorder in brain signals acquired in different modalities, such as electroencephalography, and magnetoencephalography.

Acknowledgments

The authors thank V. Aguiar-Pulido and P. López-Pérez for helpful suggestions. The work was supported by grants from Ministry of Science and Innovation of Spain with European Union, FEDER (TIN2009—07707), Xunta de Galicia, Spain (REGICC 2009/58, 08SIN010105PR, 10SIN105004PR), and CYTED (Ibero-NBIC Network 209RT0366).

References

- [1] L. Pasti, A. Volterra, T. Pozzan, and G. Carmignoto, “Intracellular calcium oscillations in astrocytes: a highly plastic, bidirectional form of communication between neurons and astrocytes *in situ*,” *Journal of Neuroscience*, vol. 17, no. 20, pp. 7817–7830, 1997.
- [2] A. Araque, V. Parpura, R. P. Sanzgiri, and P. G. Haydon, “Tripartite synapses: Glia, the unacknowledged partner,” *Trends in Neurosciences*, vol. 22, no. 5, pp. 208–215, 1999.
- [3] A. Araque, G. Carmignoto, and P. G. Haydon, “Dynamic signaling between astrocytes and neurons,” *Annual Review of Physiology*, vol. 63, pp. 795–813, 2001.
- [4] G. Perea and A. Araque, “Communication between astrocytes and neurons: a complex language,” *Journal of Physiology Paris*, vol. 96, no. 3-4, pp. 199–207, 2002.
- [5] G. Perea and A. Araque, “Properties of synaptically evoked astrocyte calcium signal reveal synaptic information processing by astrocytes,” *Journal of Neuroscience*, vol. 25, no. 9, pp. 2192–2203, 2005.
- [6] G. Perea and A. Araque, “Astrocytes potentiate transmitter release at single hippocampal synapses,” *Science*, vol. 317, no. 5841, pp. 1083–1086, 2007.
- [7] G. Perea, M. Navarrete, and A. Araque, “Tripartite synapses: astrocytes process and control synaptic information,” *Trends in Neurosciences*, vol. 32, no. 8, pp. 421–431, 2009.
- [8] G. Perea and A. Araque, “GLIA modulates synaptic transmission,” *Brain Research Reviews*, vol. 63, no. 1-2, pp. 93–102, 2010.
- [9] A. B. Porto-Pazos, N. Veiguela, P. Mesejo et al., “Artificial astrocytes improve neural network performance,” *PLoS ONE*, vol. 6, no. 4, Article ID e19109, 2011.
- [10] A. Porto, A. Araque, J. Rabuñal, J. Dorado, and A. Pazos, “A new hybrid evolutionary mechanism based on unsupervised learning for Connectionist Systems,” *Neurocomputing*, vol. 70, no. 16–18, pp. 2799–2808, 2007.
- [11] C. Ikuta, Y. Uwate, and Y. Nishio, “Chaos glial network connected to multi-layer perceptron for solving two-spiral problem,” in *Proceedings of the IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems (ISCAS ’10)*, pp. 1360–1363, June 2010.
- [12] C. Ikuta, Y. Uwate, and Y. Nishio, “Performance and features of multi-layer perceptron with impulse glial network,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN ’11)*, pp. 2536–2541, San Jose, Calif, USA, July–August 2011.
- [13] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [14] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 8, pp. 2554–2558, 1982.
- [15] W. Duch and N. Jankowski, “Survey of neural transfer functions,” *Neural Computing Surveys*, vol. 2, pp. 163–212, 1999.
- [16] A. Porto, *Modelos Computacionales para optimizar el Aprendizaje y el Procesamiento de la Información en Sistemas Adaptativos: Redes Neurogliales Artificiales (RR.NG.AA)*, Ph.D. thesis, Universidade da Coruña, 2004.
- [17] A. Porto, A. Pazos, and A. Araque, “Artificial neural networks based on brain circuits behaviour and genetic algorithms,” in *Proceedings of the 8th International Workshop on Artificial Neural Networks (IWANN ’05)*, vol. 3512 of *Lecture Notes in Computer Science*, pp. 99–106, June 2005.

- [18] A. Alvarellos, N. Veiguela, C. R. Munteanu, J. Dorado, A. Pazos, and A. B. Porto-Pazos, “The ability of MEAs containing cultured neuroglial networks to process information,” *Current Bioinformatics*, vol. 6, no. 2, pp. 199–214, 2011.
- [19] P. E. Latham, B. J. Richmond, P. G. Nelson, and S. Nirenberg, “Intrinsic dynamics in neuronal networks. I. Theory,” *Journal of Neurophysiology*, vol. 83, no. 2, pp. 808–827, 2000.
- [20] P. E. Latham, B. J. Richmond, S. Nirenberg, and P. G. Nelson, “Intrinsic dynamics in neuronal networks. II. Experiment,” *Journal of Neurophysiology*, vol. 83, no. 2, pp. 828–835, 2000.
- [21] G. L. Zhang, H. H. Lin, D. B. Keskin, E. L. Reinherz, and V. Brusic, “Dana-Farber repository for machine learning in immunology,” *Journal of Immunological Methods*, 2011.
- [22] J. Rabuñal, *Entrenamiento de Redes de Neuronas Artificiales con Algoritmos Genéticos*, Ph.D. thesis, Departamento de Computación, Facultad de Informática, Universidade da Coruña, 1998.
- [23] J. R. Rabuñal, J. Dorado, A. Pazos, J. Pereira, and D. Rivero, “A new approach to the extraction of ANN rules and to their generalization capacity through GP,” *Neural Computation*, vol. 16, no. 7, pp. 1483–1523, 2004.
- [24] A. Araque, G. Carmignoto, and P. G. Haydon, “Dynamic signaling between astrocytes and neurons,” *Annual Review of Physiology*, vol. 63, pp. 795–813, 2001.
- [25] A. Martínez and J. Goddard, “Definición de una red neuronal para clasificación por medio de un programa evolutivo,” *Revista Mexicana de Ingeniería Biomédica*, vol. 22, no. 1, pp. 4–11, 2001.

Research Article

Source Activity Correlation Effects on LCMV Beamformers in a Realistic Measurement Environment

Paolo Belardinelli,¹ Erick Ortiz,¹ and Christoph Braun^{1,2,3}

¹ MEG Center, University of Tübingen, Otfried Mueller Street 47, 72076 Tübingen, Germany

² CIMEC, Center of Mind/Brain Sciences, University of Trento, Via Delle Regole 101, 38123 Mattarello, Italy

³ DiSCoF, Department of Cognitive and Educational Sciences, University of Trento, Corso Bettini no. 31, 38068 Rovereto, Italy

Correspondence should be addressed to Paolo Belardinelli, paolo.belardinelli@gmail.com

Received 17 November 2011; Revised 1 February 2012; Accepted 9 February 2012

Academic Editor: Luca Faes

Copyright © 2012 Paolo Belardinelli et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In EEG and MEG studies on brain functional connectivity and source interactions can be performed at sensor or source level. Beamformers are well-established source-localization tools for MEG/EEG signals, being employed in source connectivity studies both in time and frequency domain. However, it has been demonstrated that beamformers suffer from a localization bias due to correlation between source time courses. This phenomenon has been ascertained by means of theoretical proofs and simulations. Nonetheless, the impact of correlated sources on localization outputs with real data has been disputed for a long time. In this paper, by means of a phantom, we address the correlation issue in a realistic MEG environment. Localization performances in the presence of simultaneously active sources are studied as a function of correlation degree and distance between sources. A linear constrained minimum variance (LCMV) beamformer is applied to the oscillating signals generated by the current dipoles within the phantom. Results show that high correlation affects mostly dipoles placed at small distances (1, 5 centimeters). In this case the sources merge. If the dipoles lie 3 centimeters apart, the beamformer localization detects attenuated power amplitudes and blurred sources as the correlation level raises.

1. Introduction

MEG and EEG provide noninvasive imaging over the whole brain with an excellent temporal resolution. The high temporal resolution qualifies these methods for the study of functional connectivity based on correlated activity. Over the last three decades, several algorithms have been developed for brain source localization [1]. Besides recent Bayesian approaches which employ iterative (and therefore computationally demanding) inversion schemes [2–4], beamformers [5, 6] still hold as one of the most reliable inversion schemes for source localization both in time and frequency domain [7–15]. Beamformers are data-dependent spatial filters. In order to pass from sensor signals to brain source activity, they employ filters which rely on the signal covariance (time domain) or the cross-spectrum density matrix (frequency domain). Moreover, beamformers assume

uncorrelated source timecourses. The covariance estimation is therefore forced to be diagonal. This may induce a bias both on location and intensity of the detected sources [16]. Some findings have shown that this assumption produces no evident bias with certain data sets. On the other hand, other studies have demonstrated that it may induce relevant biases when the level of correlation between sources and the signal-to-noise Ratio are high [17, 18]. A dual-core beamformer that takes into account the correlation effects between two sources has been implemented by Brookes et al. [19]. However, this modified beamformer implies the use of *a priori* information which is not always available. The correlation effects are particularly disruptive when analyzing brain-induced or spontaneous activity in the frequency domain. In fact, some possible remedies for modeling correlation effects have been proposed for the study of evoked potential in the time domain [20, 21]. However, these approaches based on

Bayesian theory call for a strong assumption about what is “signal” and what is “noise/spontaneous activity” within the time span of our data. This is clearly not possible in presence of data generated by spontaneous or induced brain activity.

In this paper, we have set true current dipoles within a phantom and measured the MEG signals generated by sources with various levels of mutual correlation located at different depths and mutual distances. The goal of the paper was to elucidate how correlation affects beamformer results in a realistic measurement environment. For this aim, we have localized the oscillating sources both in time and frequency domain by means of a linear constrained minimum variance (LCMV) beamformer. Since the localization results appeared extremely similar, we focused on the frequency domain which is a power implementation of dynamic imaging of coherent sources (DICS) [10].

2. Materials and Methods

2.1. Forward Solution

2.1.1. Phantom Description. In this experiment the CTF phantom model PN900-0017 was employed. This phantom consists of a 65 mm acrylic sphere, filled with saline water at 0.8% concentration, and based on an empty vertical acrylic tube. This tool emulates the brain volume conductor and the conductive medium around it. The brain current dipoles are simulated by a twisted pair of isolated wires, with open ends, encased in a glass tube. These tubes enter the sphere from the inferior part through a grid of holes that allows the dipoles to be located at different positions on the horizontal plane.

2.1.2. Head Coordinate System. To define the head coordinate system, three coils are vertically located at standard positions on the surface of the phantom sphere. The three locations loosely correspond to the fiducial points commonly used with human subjects: nasion (NAS), left periauricular point (LPA), and right periauricular point (RPA). The origin is considered the midpoint between RPA and LPA. All of the three points lie on the central horizontal plane of the sphere. The X axis points directly to NAS, and the Y axis is orthogonal to it, pointing approximately towards LPA. The Z axis is orthogonal to the XY plane, pointing at the top of the sphere.

In order to define the head localization, the three small coils (NAS, LPA, and RPA) generate a magnetic dipole signal. The center of the magnetic dipole coil is 50 mm above the plane containing the center of all three head localization coils. The dipole coil is an 11-turn single-layer air core solenoid wound on a 1.6 mm diameter mandrill. The magnetic field generated by a coil is different from the magnetic field generated by a current dipole. As a result, a different localization calculation is used for the magnetic dipole phantom.

2.1.3. Dipole Locations. Two different dipole configurations were considered: one with two parallel dipoles (#1 and #2)

placed on the same coronal plane in the two different hemispheres (Figure 1). The dipoles had different distances from the surface (2 and 3 cm, resp.). The distance between the dipoles was 3 centimeters. In the second configuration the two dipoles (#1 and #3) were placed in the left hemisphere at a slightly height (2 and 0.5 cm from the surface, resp.), with dipole #3 more external with respect to #1. Their distance in this case was 1.5 centimeters.

2.1.4. Simulated Electric Signals. The dipoles’ electric signals were divided into 200 epochs (100 epochs with oscillating dipoles, 100 with inactive ones) with a length of 0.8 seconds. The time courses were generated synthetically from a frequency distribution centered at 10 Hz, controlled for the desired correlation levels, and then commuted into electric signals by means of a digital to analog converter (DAC).

A computer with a DAC board Adlink ACL-6126 was used, with one independent channel per dipole, with alternating current output in a range of $\pm 5 \mu\text{A}$. Since we placed one dipole (#3, Figure 1) on a location which can be roughly associated to the sensorimotor region, a typical range of electrical activity for somatosensory responses was used [22]. Our DAC sampling rate was 200 Hz, the MEG sampling rate was 293 Hz. The DAC board employed for our experiment has bipolar outputs, with a common ground. Thus, while the voltage output was bipolar, the pair was not allowed to float.

Under the conditions mentioned above, the ideal dipole current drivers are optically isolated current sources; however, in the absence of this option, the best solution was (1) to use a DAC with differential outputs (range: $\pm 5 \text{ V}$) with a 1 Mohm resistor attached to both the positive and negative outputs, and (2) to ensure adequate separation (1.5 or 3 cm) between the dipole pairs, compared to the dipole length (3 mm). Part (1) of the solution ensures a known and matched value for the current in each cable of the dipole pair, and part (2) keeps the current between the pairs one order of magnitude lower than in each dipole.

2.1.5. Synthetic Time-Course Generation. For each time sample, an instantaneous frequency was drawn from a Gaussian distribution centered on a 10 Hz frequency: $N(10 \text{ Hz}, 3 \text{ Hz})$. The final time course consisted of the cumulative sum of such instantaneous frequencies, with a random initial phase. The dipole time courses were controlled either for low (0.15 ± 0.05) medium (0.55 ± 0.05) or high (0.95 ± 0.05) correlation (Figure 2).

In Figure 3 we show a plot of sensor data in the time domain and a sensor-power plot of Fourier transformed data at the frequency of interest of 10 Hz.

2.2. Spatial Filters. Since the focus of our work is to quantify the influence of correlation in the localization accuracy of dipoles with known locations, the choice of source mapping strategy was a voxel-wise spatial filter, named beamformer. Parameters for these spatial filters depend both on the forward model chosen (source distribution and a volume conductor model) and the data.

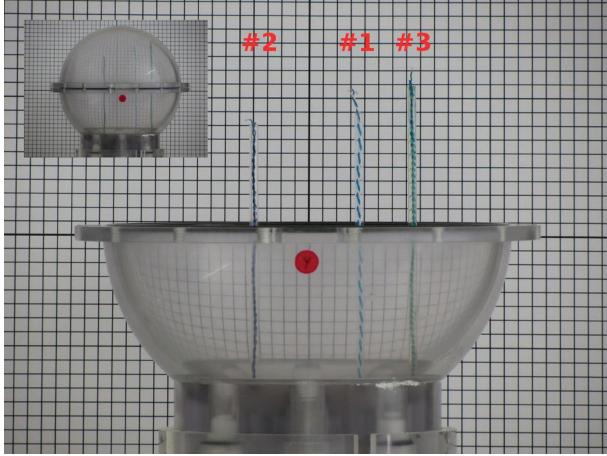


FIGURE 1: Setting of the three dipoles within the coverless phantom. Squares have a 5 mm side. In the left upper corner, the complete setup is shown.

2.2.1. Forward Model. A three-dimensional grid with 5 mm step was employed for source analysis, bounded by a sphere of 65 mm radius centered at the origin. At each grid point (voxel), the full rank-3 leadfield is calculated, and subsequently reduced to rank-2, since in a spherical conductor model the radial component is zero [23]. The volume conductor model is a sphere with a radius of 7 cm.

2.2.2. Linear Constrained Minimum Variance (LCMV) Beamformer. Linear Constrained Minimum Variance beamformers are widely employed both in time and frequency domain [9, 10]. As a first step, the 100 active epochs and the 100 inactive ones are Fourier transformed and then averaged. DICS source power mapping procedure was applied to the data to both averages. Then, the output of the silent average was used as a baseline. The active average was contrasted to the baseline.

LCMV and DICS consist in the following procedure: a filter matrix \mathbf{A} is employed in a linear transformation from the sensor level to the brain space. \mathbf{A} filters the source activity (in a given frequency band or time window) at the i th voxel (grid point) with unit gain while suppressing contribution from the other voxels. The filter depends on the data by means of the covariance $\mathbf{C}(t)$ (LCMV, time domain) or the cross spectral density (CSD) $\mathbf{C}(f)$ (DICS, frequency domain). Since the two domains are dual, we will define both matrices as \mathbf{C} in the following description of the filtering procedure. The minimization problem which yields \mathbf{A} has the following solution:

$$\mathbf{A}_i = \left(\mathbf{L}_i^T \mathbf{C}_r \mathbf{L}_i \right)^{-1} \mathbf{L}_i^T \mathbf{C}_r^{-1}, \quad (1)$$

where $\mathbf{C}_r = \mathbf{C} + \alpha \mathbf{I}$ and α is a regularization parameter. In our case $\alpha = 15\%$, the time window of interest t was the entire epoch and $f = 10 \pm 3$ Hz so that most of the signal information content was considered. \mathbf{L} is the leadfield matrix. The columns of \mathbf{L} contain the solution of the forward problem for three orthogonal unit current

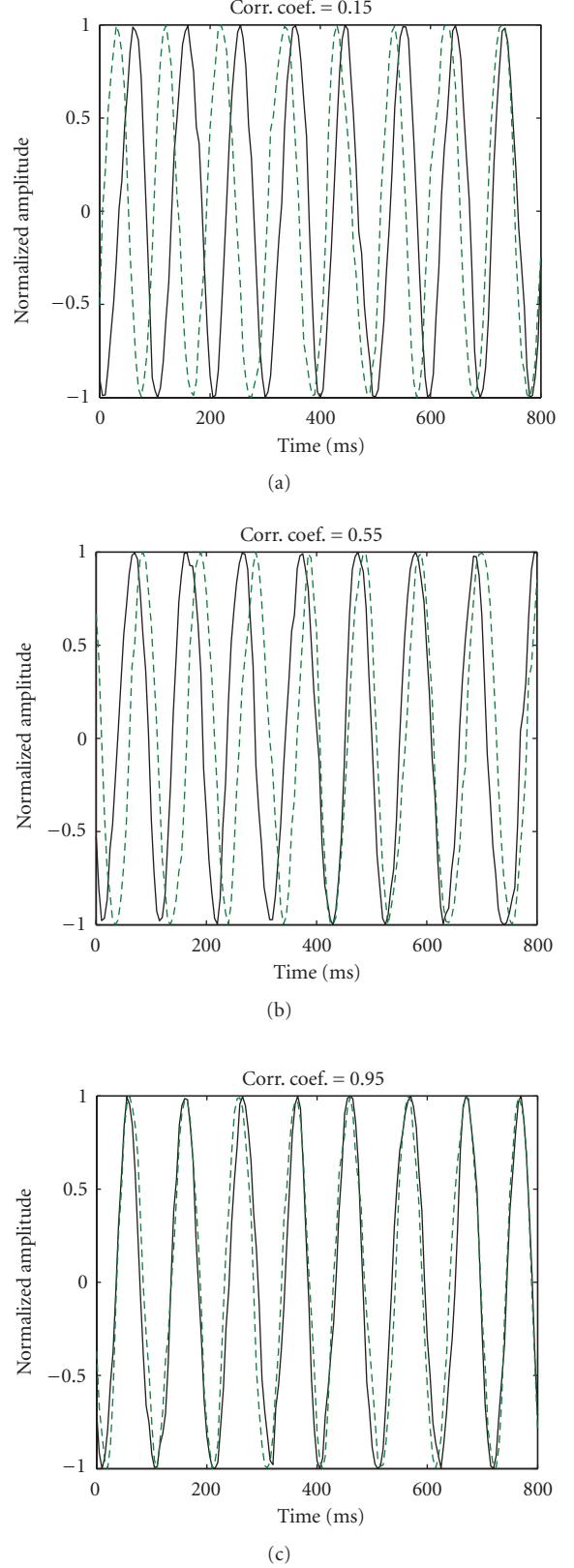


FIGURE 2: Time-courses for three different levels of correlation. Dashed and solid curves represent the dipole activities on the two source locations. Due to the addition of Gaussian noise, the signals diverge slightly from sine waves. Their phase synchrony grows as correlation increases.

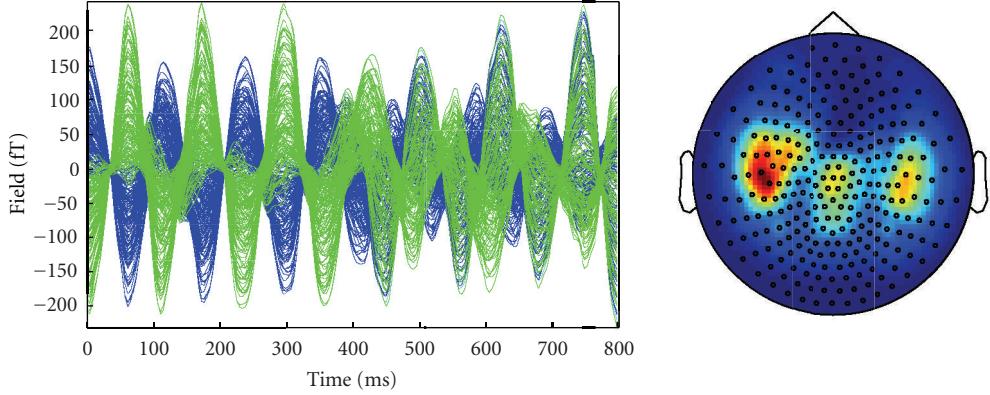


FIGURE 3: Sensor level plots generated by dipoles in positions #1 and #2, at 15% correlation. (a) signal time courses; (b) topographic plot of signal power at 10 Hz.

dipoles placed at the i th voxel. However, since the dipole radial component is silent, the leadfield rank at each site is 2. In a spherical conductor, the tangential eigenvectors span the space containing all possible source orientations that can be detected with MEG. The quantity $(\mathbf{L}_i^T \mathbf{C}_r \mathbf{L}_i)^{-1}$ is often referred to as the beamformer gain factor.

The source cross-power estimates between the two dipole components at the i th voxel are given by:

$$\mathbf{P}_i = \mathbf{A}_i \mathbf{C}_r \mathbf{A}_i^{*T}. \quad (2)$$

If the condition $\lambda_1 \gg \lambda_2$ holds for the singular values of \mathbf{P} , the source can be considered to have a fixed orientation. Otherwise, the power estimate can be obtained by computing the trace of the \mathbf{P} matrix.

In this paper, the implementations of LCMV and DICS present in the Fieldtrip package were employed (<http://fieldtrip.fcdonders.nl/>).

3. Results

3.1. Localization Results. Since the localization results of LCMV and DICS appear extremely similar (as one should expect), we will focus on DICS results. The power mappings of dipoles oscillating with low, medium, and high correlation are shown in Figure 4. In presence of a low correlation level, we obtain a good localization result for both couples of dipoles (absolute maxima on dipole sites, Figure 4(a) (couple #1 + #2) and Figure 4(d) (couple #1 + #3)). Performances decrease only slightly for a correlation level of 0.55 (Figures 4(b) and 4(e)). The localizations are marginally more blurred than in the previous case, and the relative power is faintly reduced. In the case of the close dipoles #1 and #3 one could get the deceptive idea that a 55%

source correlation level provides for better results than 15% (Figures 4(d) and 4(e), lower panel, XY plane). This is only because source #3 is detected as more blurred, and its presence is perceivable in the lower XY plane where the absolute maximum of source #1 is found. In the case of high correlation, the two sources are still recognizable for the couple #1 + #2 (Figure 4(c)) whereas couple #1 + #3 is detected as a single source (Figure 4(f)).

3.2. Detected Power Levels at Sites of Interest. Here we focus on power levels at four locations of interest in order to discern the correlation effects in a realistic environment (Figure 5). The three dipole sites are considered (deep blue, light blue, and yellow). In addition, based on the previous results, we focus on the median point between dipole #1 and #3 (red bar). Due to the spatial proximity, the dipole couple #1 + #3 tends to merge when the correlation degree increases. This does not happen in the case of the couple #1 + #2. In this case the distance always prevents the detection of these dipoles as one merged source. Differently from the couple #1 + #3, the dipoles just attenuate each other's power in the localization process.

For dipoles #1 and #2, the absolute maxima of source power detection are always correct. No relevant contribution to the power mapping is coming from the site between dipoles #1 and #3 as well as from dipole #3 at correlation levels lower than 95%, proving DICS's remarkable spatial accuracy. For high correlation the dipole sites show only 10 to 15% of the power detected in the low correlation simulation. The power level on the other two sites is not irrelevant anymore, if compared to the actual dipole sites.

For dipoles #1 and #3, due to the limited distance, 50% of power of the external source (dipole #3) is not detected at

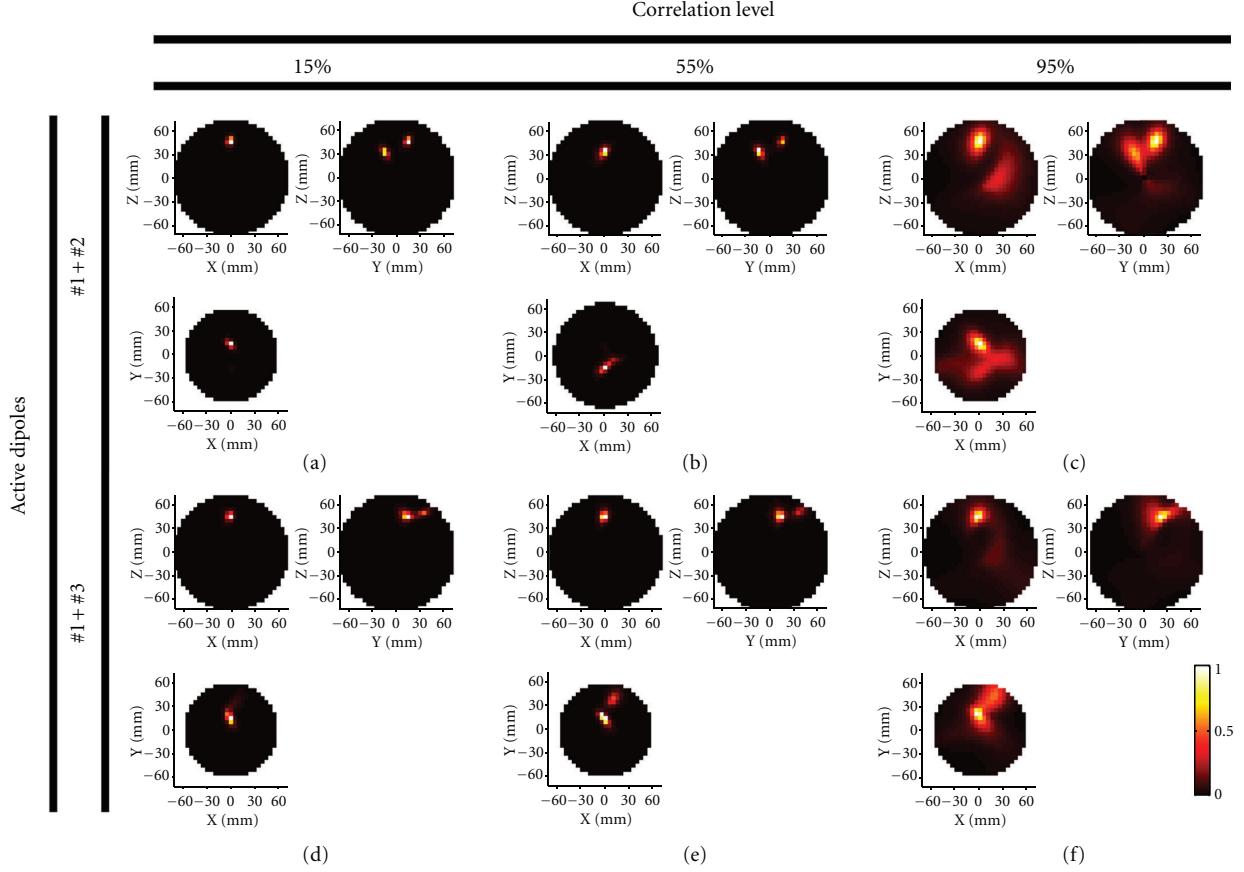


FIGURE 4: Localization results of the two different dipole couples: couple #1 and #2 (left and right, 3 cm distance, in the higher panel) and couple #1 and #3 (both left, 1.5 cm distance, in the lower panel).

low correlation. At a correlation degree of 55%, the intensity of dipole #3 is 30% of dipole #1, but the low level of power in the site between the dipoles shows that the segmentation between the two detected sources is still acceptable. Only at 95% correlation level, the two sources are not detectable anymore. The power level on the site between #1 and #3 is higher than in site of dipole #3.

Finally, the localization error (LE) for each condition and each source was calculated. Results in Table 1 show that the error is in the order of one voxel up to a correlation level of 55%.

4. Discussion

Connectivity studies by means of whole-head non invasive techniques as MEG and EEG are essential to get an insight on brain functional networks. However, the choice of the best algorithm for the network detection is not a trivial task [24]. Among several approaches [25–27], LCMV still stands as one of the most powerful schemes for detection of functional connectivity [11, 28–31]. However, LCMV beamformers are prone to correlation bias effects. Since these effects have been univocally shown only in a PC simulation context, we simulated sources in a human phantom in order to estimate how correlation affects the detection of

different network hubs in a real MEG measurement involving every kind of environmental noise. Our findings show that, with data recorded in a real MEG shielded cabin, LCMV suffers from a relevant bias only when the correlation degree between the sources is extremely high. The 55% level is already a remarkably elevated degree of correlation for the detected activity of brain functional networks [32, 33]. It is worth noting that in noninvasive studies the detected time courses of network hub activities in the human brain are reconstructed with some kind of inverse scheme. This implies a certain degree of inaccuracy. Therefore the original correlation of source time-courses could be relatively higher.

It should be noted that this study has different limitations. The number of simultaneously active sources was limited to two in order to study correlation effects in a real recording environment. More complex configurations should be studied in the future. Papadelis and colleagues [22], using a spheroid phantom similar to the one employed in the present study, estimated the localization accuracy of a synthetic-aperture magnetometry (SAM) beamformer [34] when three dipolar sources were simultaneously active. In this case the beamformer failed to detect the third source within 30 mm of the other two. Differently from that study, our results do not show a performance decrease for the deepest source (dipole #2) with respect to the other ones.

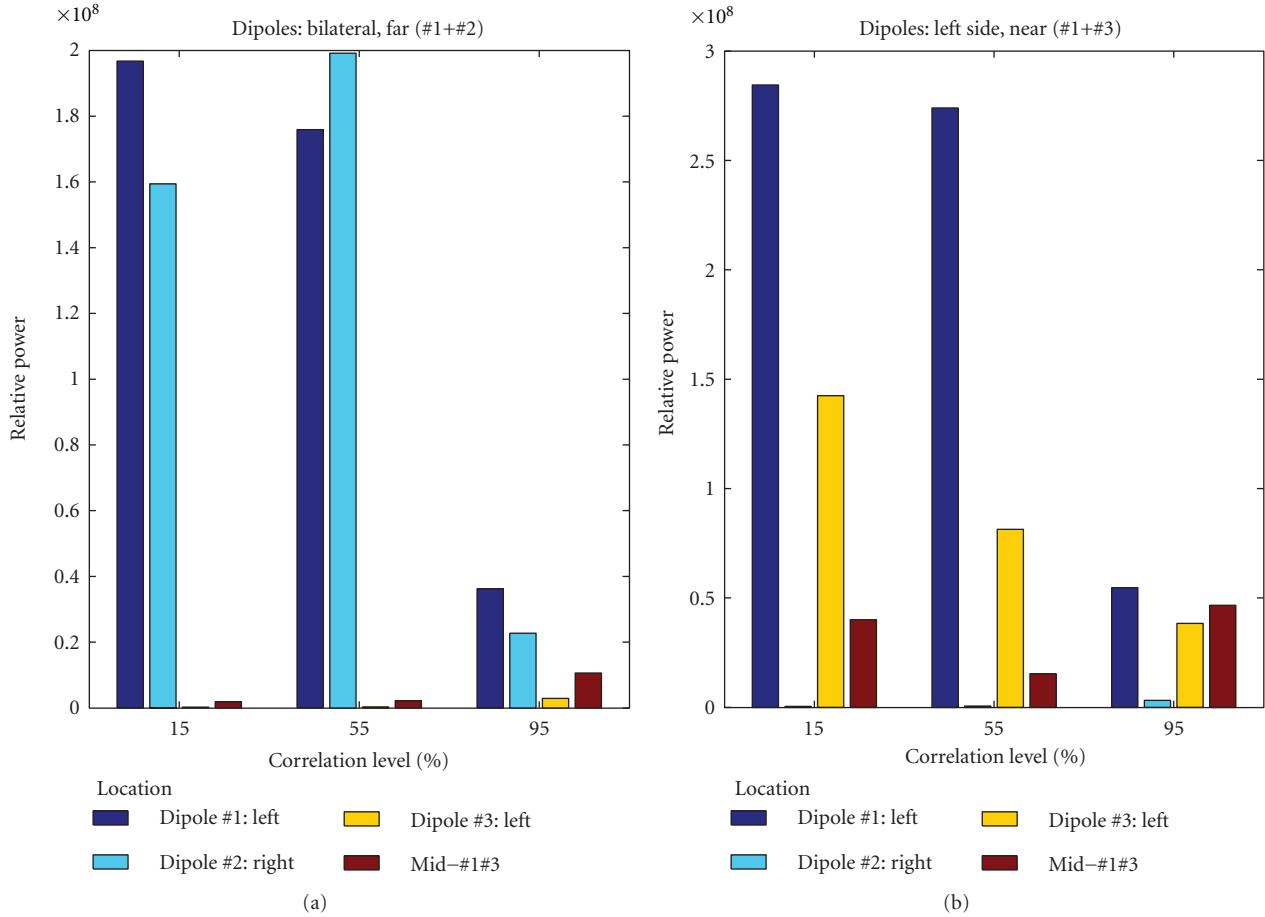


FIGURE 5: Plots of relative power normalized to the baseline obtained by means of DICS at four sites of interest: location of dipole #1 (deep blue), dipole #2 (light blue), and dipole #3 (yellow) and the median point between dipole #1 and #3 (red). The high relative values are a symptom of the good performance of the spatial filter.

TABLE 1: Localization errors (LEs) in the LCMV outputs. Results are acceptable up until a 55% level of correlation. To our surprise, results for source #1 and source #3 (nearby sources) at 55% were slightly better than findings on the same couple of sources with 15% correlation as well as with far sources #1 and #2 simultaneously active at same correlation level (55%). Differences were always in the order of 1 voxel (5 mm). For high correlation between sources (95%), a more relevant error is detectable for nearby sources whereas only a single local maximum is retrievable in the area between the two sources for the far ones.

Correlation level	LE source #1- source #2 (mm)	LE source #1- source #3 (mm)
15%	S1, 5; S2, 5	S1, 5; S3, 10
55%	S1, 11; S2, 9	S1, 5; S3, 7
95%	S1, 15; S2, 12	Single peaks not detectable

Our data suggest that MEG should be able to localize 3 cm deep sources under the condition of a sufficient number of trials (i.e., sufficiently high SNR). Furthermore, the use of a spheroid phantom is not optimal for a realistic MEG simulation. A more realistic approach should consider different dipole orientations and time shifts of sources as well as different time-course envelopes. A possible further step in the direction of realistic conditions would be the construction of a human-shaped phantom with different compartments for

brain, cerebrospinal fluid (CSF), skull, and scalp. This could potentially yield an increase in the accuracy as shown by the MEG results in [35]. The significance of precise conductivity values for MEG, where only one volume conductor is used, has been downplayed by several studies [35–38]; the shape of the volume conductor plays a more relevant role. The simplest, first-order head model is a sphere; a higher-order model could potentially map more subtleties of the head anatomy. Leahy and colleagues show that with

a realistic skull phantom and a corresponding BEM head model, the MEG localization errors are comparable with the potential registration errors. Furthermore, they found out that the localization errors induced by a locally fitted sphere are slightly larger than those generated by a BEM model (3.03 mm *versus* 3.47 mm in the localization error; 6.8° *versus* 7.7° in the tangential component error). These findings suggest that, while generating a subject-specific volume conductor is probably adding excessive complexity, a standard BEM model scaled to the subject (i.e., a global rescaling transformation with 7 parameters: 6 parameters for rotation and translation and 1 for scaling) could be a fair compromise. This information can be extracted just from the fiducial points of the subject. We performed a brief comparison between a spherical and a “Nolte” model [36]. For the sphere, only rank-2 leadfields were used, for the reasons mentioned in the methods section; this does not necessarily apply to the “Nolte” model. The 7-parameter transformation was then applied to a boundary element method (BEM) model generated from the MRI image of a subject’s head, and the ratio between rank-3 and rank-2 leadfields was calculated for each position on a regular grid with 5 mm resolution. The median ratio resulted smaller than 0.1%, indicating that most information input in the model is nondescriptive. Only 3.7% of the sites (435 out of 11654) had a ratio larger than 1%, and less than 0.2% (20 sites) crossed a 5% threshold. This finding suggests that the use of a realistic phantom and a corresponding accurate model appears necessary only in particularly elaborated simulations.

The approach with real current dipoles employed in this study can be further extended in the future to investigate human brain functional networks (resting state, acoustic, sensorimotor, etc.) comparing real and simulated (non-synthetic) data with the limitations described above. A comparison between human networks and simulated ones at different correlation levels can provide new insights both on the physiological meaning of such human networks and on beamformer limitations as a detection tool for connectivity studies. In fact, the LCMV beamformer can be employed in two simple ways in order to access brain connectivity: 1. In the time domain in order to reconstruct the source time-courses from different areas and in a second step to study their connectivity by means of different algorithms [11, 39–41]. Another possible extension of the present study is the use of DICS not just for power but also for coherence mapping. In this last case, instead of calculating the source power estimates at the i th voxel by means of (2), the cross spectrum and coherence estimates between the tangential dipoles at the voxels placed at \mathbf{r}_1 and \mathbf{r}_2 can be estimated for every brain voxel [10].

Acknowledgment

The present work is supported by an award from the Center for Integrated Neuroscience (CIN) of Tübingen (Germany) (Pool Project no. 2010-17).

References

- [1] S. Baillet, J. C. Mosher, and R. M. Leahy, “Electromagnetic brain mapping,” *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14–30, 2001.
- [2] K. Friston, L. Harrison, J. Daunizeau et al., “Multiple sparse priors for the M/EEG inverse problem,” *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, 2008.
- [3] M. A. Sato, T. Yoshioka, S. Kajihara et al., “Hierarchical Bayesian estimation for MEG inverse problem,” *NeuroImage*, vol. 23, no. 3, pp. 806–826, 2004.
- [4] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, “Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG,” *NeuroImage*, vol. 49, no. 1, pp. 641–655, 2010.
- [5] M. D. Zoltowski, “On the performance analysis of the MVDR beamformer in the presence of correlated interference,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 6, pp. 945–947, 1988.
- [6] B. D. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, “Localization of brain electrical activity via linearly constrained minimum variance spatial filtering,” *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 867–880, 1997.
- [7] G. R. Barnes and A. Hillebrand, “Statistical flattening of MEG beamformer images,” *Human Brain Mapping*, vol. 18, no. 1, pp. 1–12, 2003.
- [8] P. Belardinelli, L. Ciancetta, M. Staudt et al., “Cerebromuscular and cerebro-cerebral coherence in patients with pre- and perinatally acquired unilateral brain lesions,” *NeuroImage*, vol. 37, no. 4, pp. 1301–1314, 2007.
- [9] D. Cheyne, A. C. Bostan, W. Gaetz, and E. W. Pang, “Event-related beamforming: a robust method for presurgical functional mapping using MEG,” *Clinical Neurophysiology*, vol. 118, no. 8, pp. 1691–1704, 2007.
- [10] J. Gross, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, and R. Salmelin, “Dynamic imaging of coherent sources: studying neural interactions in the human brain,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 2, pp. 694–699, 2001.
- [11] J. Gross, L. Timmermann, J. Kujala et al., “The neural basis of intermittent motor control in humans,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 4, pp. 2299–2302, 2002.
- [12] S. D. Hall, I. E. Holliday, A. Hillebrand et al., “The missing link: analogous human and primate cortical gamma oscillations,” *NeuroImage*, vol. 26, no. 1, pp. 13–17, 2005.
- [13] N. Hoogenboom, J. M. Schoffelen, R. Oostenveld, L. M. Parkes, and P. Fries, “Localizing human visual gamma-band activity in frequency, time and space,” *NeuroImage*, vol. 29, no. 3, pp. 764–773, 2006.
- [14] H. B. Hui, D. Pantazis, S. L. Bressler, and R. M. Leahy, “Identifying true cortical interactions in MEG using the nulling beamformer,” *NeuroImage*, vol. 49, no. 4, pp. 3161–3174, 2010.
- [15] V. Litvak, A. Eusebio, A. Jha et al., “Optimized beamforming for simultaneous MEG and intracranial local field potential recordings in deep brain stimulation patients,” *NeuroImage*, vol. 50, no. 4, pp. 1578–1588, 2010.
- [16] K. Sekihara, S. S. Nagarajan, D. Poeppel, and A. Marantz, “Performance of an MEG adaptive-beamformer technique in the presence of correlated neural activities: effects on signal intensity and time-course estimates,” *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 12, pp. 1534–1546, 2002.

- [17] K. Sekihara, S. S. Nagarajan, D. Poeppel, and A. Marantz, "Performance of an MEG adaptive-beamformer source reconstruction technique in the presence of additive low-rank interference," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 1, pp. 90–99, 2004.
- [18] D. Wipf and S. Nagarajan, "Beamforming using the relevance vector machine," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, pp. 1023–1030, June 2007.
- [19] M. J. Brookes, C. M. Stevenson, G. R. Barnes et al., "Beamformer reconstruction of correlated sources using a modified source model," *NeuroImage*, vol. 34, no. 4, pp. 1454–1465, 2007.
- [20] J. M. Zumer, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data," *NeuroImage*, vol. 37, no. 1, pp. 102–115, 2007.
- [21] A. Dogandzic and A. Nehorai, "Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 13–25, 2000.
- [22] C. Papadelis, V. Poghosyan, P. B. C. Fenwick, and A. A. Ioannides, "MEG's ability to localise accurately weak transient neural sources," *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1958–1970, 2009.
- [23] J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Physics in Medicine and Biology*, vol. 32, no. 1, article 11, 1987.
- [24] J. M. Schoffelen and J. Gross, "Source connectivity analysis with MEG and EEG," *Human Brain Mapping*, vol. 30, no. 6, pp. 1857–1865, 2009.
- [25] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, 2003.
- [26] K. Jerbi, J. P. Lachaux, K. N'Diaye et al., "Coherent neural representation of hand speed in humans revealed by MEG imaging," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 18, pp. 7676–7681, 2007.
- [27] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.
- [28] P. Fries, "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence," *Trends in Cognitive Sciences*, vol. 9, no. 10, pp. 474–480, 2005.
- [29] M. Liljeström, J. Kujala, O. Jensen, and R. Salmelin, "Neuromagnetic localization of rhythmic activity in the human brain: a comparison of three methods," *NeuroImage*, vol. 25, no. 3, pp. 734–745, 2005.
- [30] A. Schnitzler and J. Gross, "Normal and pathological oscillatory communication in the brain," *Nature Reviews Neuroscience*, vol. 6, no. 4, pp. 285–296, 2005.
- [31] J. M. Schoffelen, R. Oostenveld, and P. Fries, "Neuronal coherence as a mechanism of effective corticospinal interaction," *Science*, vol. 308, no. 5718, pp. 111–113, 2005.
- [32] F. De Pasquale, S. Della Penna, A. Z. Snyder et al., "Temporal dynamics of spontaneous MEG activity in brain networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 13, pp. 6040–6045, 2010.
- [33] D. Mantini, M. G. Perrucci, C. Del Gratta, G. L. Romani, and M. Corbetta, "Electrophysiological signatures of resting state networks in the human brain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 32, pp. 13170–13175, 2007.
- [34] S. E. Robinson and J. Vrba, "Functional neuroimaging by synthetic aperture magnetometry (SAM)," in *Proceedings of the 11th International Conference on Biomagnetism (Biomag '98)*, pp. 302–305, Sendai, Japan, August 1998.
- [35] R. M. Leahy, J. C. Mosher, M. E. Spencer, M. X. Huang, and J. D. Lewine, "A study of dipole localization accuracy for MEG and EEG using a human skull phantom," *Electroencephalography and Clinical Neurophysiology*, vol. 107, no. 2, pp. 159–173, 1998.
- [36] G. Nolte, "The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors," *Physics in Medicine and Biology*, vol. 48, no. 22, pp. 3637–3652, 2003.
- [37] A. M. Dale and M. I. Sereno, "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach," *Journal of Cognitive Neuroscience*, vol. 5, no. 2, pp. 162–176, 1993.
- [38] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain," *Reviews of Modern Physics*, vol. 65, no. 2, pp. 413–497, 1993.
- [39] E. Florin, J. Gross, J. Pfeifer, G. R. Fink, and L. Timmermann, "The effect of filtering on Granger causality based multivariate causality measures," *NeuroImage*, vol. 50, no. 2, pp. 577–588, 2010.
- [40] J. P. Lachaux, E. Rodriguez, M. Le Van Quyen, A. Lutz, J. Martinerie, and F. J. Varela, "Studying single-trials of phase synchronous activity in the brain," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 10, no. 10, pp. 2429–2439, 2000.
- [41] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Physical Review Letters*, vol. 76, no. 11, pp. 1804–1807, 1996.

Review Article

Inferring Functional Neural Connectivity with Phase Synchronization Analysis: A Review of Methodology

Junfeng Sun,¹ Zhijun Li,² and Shanbao Tong¹

¹Med-X Research Institute, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China

²Department of Automation, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Shanbao Tong, stong@sjtu.edu.cn

Received 25 November 2011; Accepted 31 January 2012

Academic Editor: Dimitris Kugiumtzis

Copyright © 2012 Junfeng Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Functional neural connectivity is drawing increasing attention in neuroscience research. To infer functional connectivity from observed neural signals, various methods have been proposed. Among them, phase synchronization analysis is an important and effective one which examines the relationship of instantaneous phase between neural signals but neglecting the influence of their amplitudes. In this paper, we review the advances in methodologies of phase synchronization analysis. In particular, we discuss the definitions of instantaneous phase, the indexes of phase synchronization and their significance test, the issues that may affect the detection of phase synchronization and the extensions of phase synchronization analysis. In practice, phase synchronization analysis may be affected by observational noise, insufficient samples of the signals, volume conduction, and reference in recording neural signals. We make comments and suggestions on these issues so as to better apply phase synchronization analysis to inferring functional connectivity from neural signals.

1. Introduction

Segregation and integration are two fundamental organization principles of neural systems in brain [1]. The neural organization can be investigated with functional neural connectivity in both local and global brain regions. The connectivity in local regions reflects specialized functions of local cortex, while the connectivity among spatially separated brain regions plays important roles in advanced cognitive function [2–7]. Various measures, such as phase synchronization index (PSI), mutual information, correlation coefficient, coherence, and partial directed coherence, have been applied to inferring neural connectivity among brain units at multiple temporal and spatial scales with neural signals including electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging, and local field potential [8–13]. These measures can be classified into several families such as correlation/coherence family, phase synchronization family, and Granger causality family. The measures in the same family usually yield results with strong correlation to each

other and thus could provide little additional information on neural connectivity, while some measures belonging to different families may have weak correlation to each other in inferring neural connectivity, which implies that they each could characterize interdependence of signals in different aspects [13].

Among these measures, PSI quantifies the relationship between instantaneous phases (IP, represents the rhythm of oscillation or signal wave) of coupled systems/brain units but neglects the effect of their amplitude. PSI has been demonstrated to be effective in inferring neural connectivity, especially when the connectivity is too weak to be detected by other measures [8, 14]. Note that there are other types of synchronization, such as complete synchronization and generalized synchronization [14, 15]. But most of them are defined for theoretical models of coupled oscillators and difficult to be applied to neural signal analysis. The coupled systems/units are claimed to be in phase synchronization (PS) when the difference of IPs is bounded with respect to time. PS is ubiquitous in both natural and man-made systems, such as neural oscillations [4, 16], coupled chaotic

oscillators [17], chaotic laser arrays [18], and electrochemical oscillations [19]. In addition, two metrics of PS, that is, the phase-lock interval and the lability of global synchronization, hold power law probability distributions for both simulation systems (i.e., the Ising model and the Kuramoto model) and brain networks at a broadband frequency. These results imply that the IP association among multiple units or brain regions hold the property of criticality [20].

PSI, also called phase-locking value in literature, is proposed to quantify the level of PS and has been applied to a wide range of neuroscience research [3, 5]. For example, PSI has been used to examine the alternation of cortical connectivity prior to seizures [21] and neuronal synchrony after ischemic stroke [22], to identify mental states in brain-computer interface [23, 24], to investigate cognitive dysfunctions of mental disorder [25], and to gain new strategies for clinical treatment [22, 26]. However, to get a reliable estimation of PSI from observed neural signals is not so easy, especially when the signals are with a small number of samples and/or contaminated by noise [10, 11, 27, 28]. Note that, in literature, other measures, such as phase clustering index, used different definition of phase, which is not within the framework of IP [29, 30]. In this paper, we review the methodology of PS analysis in inferring functional neural connectivity from the following aspects.

- (1) IP definition. To detect PS in a pair of signals, various IP definitions have been proposed. Most of these IP definitions are based on particular transforms, such as the Hilbert transform [17] and the wavelet transform [31]. But actually, these IP definitions can be unified into one framework which defines IP with specific filter applied to signals [32].
- (2) PSI and its significance test. We will introduce two commonly used PSIs, which are based on entropy [2, 33] and circular statistics [34, 35], respectively. We will further introduce several strategies to provide significance test for the estimated PSI.
- (3) Practical problems in PS detection. The estimation of PSI is affected by observational noise, volume conduction, the number of samples in observed signals, reference, and other factors. We will summarize the advances on these issues and give suggestions for a better PS detection.
- (4) Extensions of PS analysis. Lots of extensions/variations of PS analysis have been proposed for multitrial signals and multivariate signals. We will give a brief overview of these methods and some comments to their applications in neural signal analysis.

2. Definition of Instantaneous Phase

Before estimating the level of PS, the IP of each signal should be defined (Figure 1). The most basic definition of IP is based on the Hilbert transform [14, 17, 36], which can be directly applied to coherent signals (i.e., narrow band signals with one prominent spectral component). But, for noncoherent data such as raw neural signals, this IP

definition may yield negative instantaneous frequency (IF, defined as the derivative of IP with respect to time), which is physically meaningless [37–39]. One way is to process noncoherent signal with a specific narrow bandpass filter. An alternative way is to define IP based on other transforms such as the wavelet transform [18, 19, 31, 40]. PS detection based on these IP definitions has been compared numerically with both simulation data and experimental signals, yielding similar connectivity [10, 11]. In addition, an analytical study on these IP definitions has unified them into one framework which defines IP with a specific bandpass filter applied [32]. In this section, we will briefly introduce issues on IP definition based on results in [32].

2.1. Definitions of Instantaneous Phase

2.1.1. IP Definition Based on the Hilbert Transform. For a given signal $s(t)$, its analytic signal is defined as

$$s^{(h)}(t) = s(t) + j\tilde{s}(t), \quad (1)$$

where

$$\tilde{s}(t) = \mathcal{H}[s(t)] = \frac{1}{\pi} \text{P.V.} \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} d\tau \quad (2)$$

is the Hilbert transform of $s(t)$ (here P.V. means that the integral is taken in the sense of the Cauchy principal value). Then the IP of $s(t)$ is defined as

$$\phi^{(h)}(t) = \arg[s^{(h)}(t)] = \arctan \frac{\tilde{s}(t)}{s(t)}. \quad (3)$$

In the frequency domain, $s^{(h)}(t)$ appears as $S^{(h)}(f) = S(f)B^{(h)}(f)$, where $S(f)$ is the Fourier transform of $s(t)$ and

$$B^{(h)}(f) = \begin{cases} 2, & \text{if } f > 0 \\ 1, & \text{if } f = 0 \\ 0, & \text{if } f < 0 \end{cases} \quad (4)$$

is the Fourier transform of $b^{(h)}(t) = \delta(t) + j(1/\pi t)$. Then in numerical implementation, the estimation of analytic signal $s^{(h)}(t)$ can be obtained by the inverse Fourier transform of $S^{(h)}(f)$. In the time domain, $s^{(h)}(t)$ can be expressed as the convolution of $s(t)$ with the complex-response filter $b^{(h)}(t)$; that is,

$$s^{(h)}(t) = s(t) * b^{(h)}(t). \quad (5)$$

2.1.2. IP Definition Based on Gaussian Filter. Another method defines IP as

$$\phi^{(g)}(t) = \arg[s^{(g)}(t)], \quad (6)$$

where

$$s^{(g)}(t) = s(t) * b^{(g)}(t) \quad (7)$$

is the convolution of $s(t)$ with a narrow-band Gaussian filter $b^{(g)}(t) = (1/\sqrt{2\pi T})e^{-t^2/(2T^2)}e^{j2\pi f_n t}$, which is shifted by the

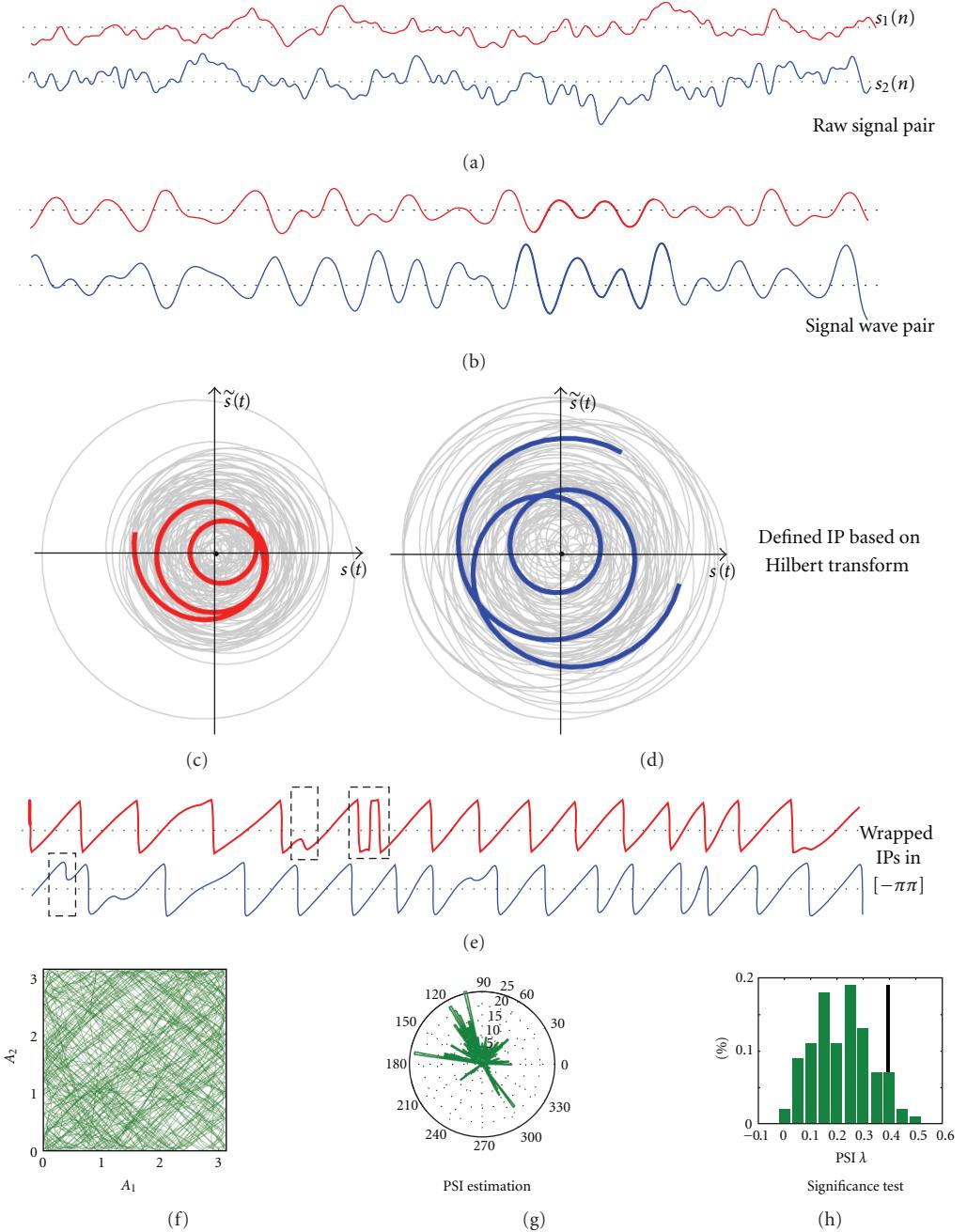


FIGURE 1: Schematic diagram of phase synchronization (PS) analysis. For broadband raw signals (a), a bandpass filter is first applied to extracting signal waves in specific frequency band (b). Then analytic signals of signal waves may be defined based on the Hilbert transform ((c) and (d)), and the argument of the analytic signals are defined as instantaneous phase (IP) of the corresponding signal waves. The IPs could be wrapped into the range $[-\pi, \pi]$ (e). In some cases as marked by dotted rectangle in (e), the estimated analytic signal $[s(t) + j\tilde{s}(t)]$ may be ill-defined due to noisy data and does not always rotate counterclockwise around origin in the complex plane, resulting in non-monotonic IP “jump” at the time when the trajectory of analytic signal crosses through the origin. Signals with too many IP “jumps” are not suitable for PS analysis. With the differences of IPs which are wrapped in the range $[-\pi, \pi]$, PS index (PSI), which quantifies the level of PS, could be estimated according to the distribution of IP difference (g). In addition, significance test could provide a significance threshold (the black bar in (h)) for estimated PSI. If the estimated PSI is greater than the threshold, then the corresponding signal wave pair is claimed to be in significant PS with a certain confidence level. For some cases, the amplitudes of analytic signals may be rather weakly correlated (f), but the corresponding PSI is with relatively large value. For the case in (f) and (g), the correlation coefficient between the amplitudes (i.e., A_1 versus A_2) of two signal waves is -0.07 , while the corresponding MPC-based PSI is 0.44 .

nominal frequency f_n [18]. In the frequency domain, $s^{(g)}(t)$ can be expressed as $S^{(g)}(f) = S(f)B^{(g)}(f)$, where $B^{(g)}(f) = e^{-2\pi^2 T^2 (f - f_n)^2}$. PS analysis based on this IP definition has successfully detected PS in coupled laser arrays, which was not revealed by IP defined with the Hilbert transform [18]. The possible reason is that the Gaussian filter extracts the components of the laser data in a particular frequency band which are in PS, while PS analysis based on the Hilbert transform does not use filter to extract the components, and thus the underlying PS level is submerged by noise and components in other bands.

2.1.3. IP Definition Based on the Wavelet Transform. With the Gabor wavelet $\psi(t) = g(t)e^{j2\pi ft}$, IP is defined as

$$\phi^{(w)}(t) = \arg[s^{(w)}(t)], \quad (8)$$

where

$$s^{(w)}(t, a) = s(t) * b^{(w)}(t), \quad (9)$$

is the convolution of $s(t)$ with $b^{(w)}(t) = f_n^{1/2}g(f_n t)e^{j2\pi f_n t}$, and $g(t) = (T^2\pi)^{-1/4}e^{-t^2/(2T^2)}$ is the envelope [31]. The difference between IP definitions based on the Gaussian filter and on the wavelet transform lies in that the filters $b^{(w)}(t)$ and $b^{(g)}(t)$ are with Gaussian windows of different amplitudes and width; that is, $b^{(w)}(t)$ is scaled by f_n . In the frequency domain, $g(t)$ appears as $G(f) = (4\pi T^2)^{1/4}e^{-2\pi^2 f^2 T^2}$, and $b^{(w)}(t)$ is $B^{(w)}(f) = f_n^{-1/2}G((f/f_n) - 1)$.

2.1.4. Unified Framework for IP Definitions. The IP definitions discussed above all can be expressed as the argument of the signal after a specific filter. In other words, these IP definitions can be unified into one common framework [32]; that is,

$$\phi^{(b)}(t) = \arg[s^{(b)}(t)], \quad (10)$$

where

$$s^{(b)}(t) = s(t) * b(t). \quad (11)$$

Here $b(t) = g(t)e^{j2\pi f_n t}$ is a filter with envelope $g(t) = (1/\sqrt{2\pi T})e^{-t^2/(2T^2)}$, nominal frequency f_n , and response duration T . The filter assures that the extracted signal wave is coherent. $s^{(b)}(t)$ is analytic in an asymptotic sense as the bandwidth of $b(t)$ is smaller than $2f_n$ [36, 41, 42]. This method is exactly the one based on a Gaussian filter when the envelope $g(t)$ is a Gaussian function. Filter $b(t)$ can be expressed as

$$B(f) = \int_{-\infty}^{\infty} b(t)e^{-j2\pi ft} dt = G(f - f_n) \quad (12)$$

in the frequency domain, where $G(f) = e^{-2\pi^2 f^2 T^2}$ is the Fourier transform of $g(t)$.

The analytic signal $s^{(b)}(t) = s(t) * b(t)$ can be interpreted as a combination of the Hilbert transform and a real

bandpass filter. Let $s^{(r)}(t) = s(t) * [g(t) \cos(2\pi f_n t)]$, where $g(t) \cos(2\pi f_n t)$ is the real part of $b(t)$. Then $s^{(r)}(t)$ appears as

$$S^{(r)}(f) = S(f) \left[\frac{1}{2}G(f + f_n) + \frac{1}{2}G(f - f_n) \right] \quad (13)$$

in the frequency domain. With (4), the analytic signal of $s^{(r)}(t)$ can be obtained by performing the inverse Fourier transform to $S^{(r)}(f)B^{(h)}(f)$; that is,

$$\begin{aligned} & \mathcal{F}^{-1}[S^{(r)}(f)B^{(h)}(f)] \\ &= \mathcal{F}^{-1}\left\{S(f)\left[\frac{1}{2}G(f + f_n) + \frac{1}{2}G(f - f_n)\right]B^{(h)(f)}\right\} \\ &= \mathcal{F}^{-1}[S(f)G(f - f_n)] \\ &= s(t) * [g(t)e^{j2\pi f_n t}] \\ &= s^{(b)}(t), \end{aligned} \quad (14)$$

where $\mathcal{F}^{-1}(\cdot)$ denotes the inverse Fourier transform operator.

Beside IP definitions within this framework, there are IP definitions proposed in other aspects [43, 44]. For example, the IP and PSI based on Rihaczek distribution are more robust to noise than the method based on the wavelet transform [43].

2.2. Constraints for IP Definition. For a real signal $s(t) = \mathcal{R}[A(t)e^{j\phi(t)}]$, its imaginary counterpart $\mathcal{I}[A(t)e^{j\phi(t)}]$ is usually unobservable, where $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ denote the real and the imaginary parts of the complex variable (\cdot) , respectively. Generally, the imaginary part is estimated from $s(t)$ by a certain operation; that is, $\mathcal{I}[A(t)e^{j\phi(t)}] = \widetilde{\mathcal{H}}[s(t)]$. Among various operators $\widetilde{\mathcal{H}}(\cdot)$ proposed, the Hilbert transform $\mathcal{H}(\cdot)$ is the only one that satisfies the following three conditions [45].

- (i) The associated amplitude $A(t)$ is continuous and differentiable.
- (ii) The IPs of signals $s(t)$ and $c \cdot s(t)$ is the same; that is, the operator should possess the property $\widetilde{\mathcal{H}}[c \cdot s(t)] = c \cdot \widetilde{\mathcal{H}}[s(t)]$, where c is a constant.
- (iii) For any constant amplitude $A > 0$, frequency $\omega > 0$, and phase ψ , the operator satisfies $\widetilde{\mathcal{H}}[A \cos(\omega t + \psi)] = A \sin(\omega t + \psi)$.

In addition, the Bedrosian theorem gives more constraints on IP definition based on the Hilbert transform [36, 42]. This theorem states that the relation

$$\mathcal{H}[l(t)h(t)] = l(t)\mathcal{H}[h(t)] \quad (15)$$

holds for a low-frequency term $l(t)$ and a high-frequency term $h(t)$ whose spectra do not overlap in the frequency domain. For real signal $s(t) = A(t) \cos \phi(t)$, $s^{(q)}(t) = A(t)e^{j\phi(t)}$ is its *quadrature model*. It seems that $\phi(t)$ in this model is a good candidate for definition of IP. However, it is difficult to estimate $A(t)$ and $\phi(t)$ from only the observable

signal $s(t)$ with this model. There is a difference between $s^{(q)}(t)$ and the analytic signal $s^{(h)}(t)$ (1). This difference approaches zero in an asymptotic sense as $A(t)$ and $\cos \phi(t)$ fulfill the Bedrosian theorem [42]. Note that the Bedrosian theorem sets a constraint to the signal, which is similar to the second condition; that is, $\widetilde{\mathcal{H}}[c \cdot s(t)] = c \cdot \widetilde{\mathcal{H}}[s(t)]$, that set on the operator. For noncoherent signal, bandpass filter $b(t)$ is usually applied, so that the extracted signal wave $s^{(b)} = s(t) * b(t)$ is coherent and fulfills the Bedrosian theorem. The effective bandwidth of filter $b(t)$ is $\Delta f = 1/(2\sqrt{2}\pi T)$, which should be less than $2f_n$ [32]. As the filter $b(t)$ becomes narrower close to delta function (i.e., $\delta(f - f_n)$) in the frequency domain, the IF ($1/2\pi$) $(d\phi^{(b)}(t)/dt)$ of the components in the pass band approaches the nominal frequency f_n in an asymptotic sense [46].

Generally, to infer functional connectivity with PS analysis, we would recommend to define IP by combining the Hilbert transform with specific bandpass filter; that is, a bandpass filter is first applied to extracting the neural signal waves in specific frequency band, and then IP is defined based on the Hilbert transform.

3. Phase Synchronization Analysis

3.1. Phase Synchronization Indexes. Let $\phi_1(t)$ and $\phi_2(t)$ denote the cumulative IP of signals observed from two coupled units, respectively. Then the signal pair is said to be in $p:q$ PS when the inequality $|p\phi_1(t) - q\phi_2(t)| < \text{const.}$ holds, where p and q are positive integers. In this paper, we focus on the case of 1:1 PS, which is defined based on samples of one trial in the time domain. Most conclusions for 1:1 PS can be easily extended to the case of $p:q$ PS [47, 48]. In neuroscience research, the estimated PSI is usually taken as one kind of functional connectivity in neural signals. Here we introduce two PSIs which have been commonly used to quantify the functional connectivity of neural signals. They are based on entropy [2, 33] and circular statistics [34, 35], respectively. Let $s(n)$ denote $s(n\Delta t)$, that is, the observation of $s(t)$ at time $n\Delta t$, and let $\hat{\phi}(n)$ denote the estimation of $\phi(t)$ at time $n\Delta t$, where Δt is the sampling interval.

(i) The entropy-based PSI is estimated by

$$\rho = \frac{(Q_{\max} - Q)}{Q_{\max}}, \quad (16)$$

where $Q = -\sum_{i=1}^K P_i \ln(P_i)$ is the entropy of distribution $P(\hat{\phi})$, $\hat{\phi}(n) = \hat{\phi}_1(n) - \hat{\phi}_2(n)$, $Q_{\max} = \ln K$, and K is the number of bins of distribution [2, 33].

(ii) The mean phase coherence (MPC) of IP difference $\varphi(t) = \phi_1(t) - \phi_2(t)$ is another commonly used PSI. It is defined as $\lambda = |E[e^{j\varphi}]|$ and can be estimated via

$$\lambda = \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{j\hat{\phi}(n)} \right|, \quad (17)$$

where $\{\hat{\phi}(n)\}_{n=0}^{N-1}$ is the estimation of $\varphi(t)$ and N is the number of samples.

These two PSIs quantify how concentrated the distribution of IP difference is. Their values are in $[0, 1]$, with $\text{PSI} = 1$ implying that the signal pair is with exact rhythm locking and $\text{PSI} = 0$ indicating no PS at all. Beside these two indexes, other measures, such as the index based on conditional probability [2] and the index based on mutual information [49], have also been applied to the quantification of relationship between IPs. More discussions on PSI can be found in [8].

3.2. Significance Tests for PSI. In inferring neural connectivity, a highly concerned question is whether the estimated PSI could indicate that the signal pair is with significant connective strength or not. This question is also concerned in brain network research, as brain network is constructed by setting edges between nodes (i.e., brain regions) with connective strength over a predefined threshold. Accumulated studies show that both functional and structural brain networks are of small-worldness with sparse edge number [28, 50–55]. The topology of brain networks are dependent on the number of edges, that is, the number of connectivity with significantly big strength. One way to decide the value of this common threshold is based on significance test for estimated PSI [8, 34, 49, 56]. Here we introduce three different strategies in providing significance threshold for PS analysis.

3.2.1. Artificial Surrogate Tests. The first strategy is based on artificial surrogate data. Surrogate methods usually produce artificial data by randomizing the concerned property but keeping as much as possible other properties of the original signal [49, 57]. Then whether the original signal possesses the concerned property could be tested with the artificial surrogate data. Various surrogate methods have been proposed [8, 57–61]. For example, surrogate method based on autoregressive model generates surrogate realizations by fitting an autoregressive model of the original signals and using independent white noise as the inputs to the model. The surrogate data generated by this method are linear stochastic processes and have the same power spectra of the original signals [57, 60]. In another study, a method named twin surrogate is proposed based on the recurrence properties. It is demonstrated to be suitable to provide significance test for PS in Rössler oscillators [61].

In [28], four different surrogate methods, which generate artificial surrogate data by shuffling the rank order, the phase spectra, the IP of original EEG signals, are compared in significance test for PS analysis. Results show that the phase-shuffled surrogate method is workable for PS analysis. This method generates surrogate data by shuffling the phase spectra of original signal but keeping the amplitude spectra unchanged. Let $\{S(k)\}_{k=0}^{N-1}$ denote the discrete Fourier transform of original signal $\{s(n)\}_{n=0}^{N-1}$; that is,

$$S(k) = \sum_{n=0}^{N-1} s(n) e^{-j2\pi kn/N}, \quad k = 0, 1, \dots, N-1. \quad (18)$$

Then a surrogate realization of $\{s(n)\}$ is generated by the inverse discrete Fourier transform of $\tilde{S}(k)$; that is,

$$\tilde{s}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{S}(k) e^{j2\pi kn/N}, \quad n = 0, 1, \dots, N-1, \quad (19)$$

where

$$\tilde{S}(k) = |S(k)| e^{j\nu(k)}, \quad (20)$$

with a uniform random sequence $\{\nu(k)\}_{k=0}^{N-1}$ [57, 58].

A $(1 - \alpha) \times 100\%$ level of significance corresponds to a probability α of a false rejection. A larger number of surrogate realizations could offer a greater power in discrimination [57]. In [28], 99 surrogate pairs are generated for each original EEG signal pair. Then if the PSI of one original signal pair is larger than the 5th biggest in all the 100 PSIs (i.e., the PSIs of the original signal pair and its 99 surrogate pairs), the original signal pair is claimed to be in PS with a 95% level of significance. Figure 2 shows the histogram of the MPC-based PSIs λ for 435 original EEG signal pairs and their phase-shuffled surrogate pairs for one subject, where one common significance threshold is estimated for all 435 original signal pairs in each case.

3.2.2. Analytical Significance Test. The second strategy is based on the distribution test of IP series [34, 56]. When the distribution of PSI is well approximated by an analytical model, a significance threshold then could be offered analytically. Empirical distributions of IP difference of coupled Rössler systems have been tested under the assumption that IP obeys specific distributions and the IPs of different samples are independent [56]. But this is applicable only for special cases, as the distribution of IP difference may vary with respect to different dynamical systems. In another study, the IP difference is formulated as an increment process under the assumption that the increments could be represented by an α -mixing process. A theoretical significance level is proposed for the MPC-based PSI. Simulation results show that the significance level is workable when the time series is with enough samples and two dependent parameters could be reliably estimated [34].

3.2.3. Surrogate Test Based on Intersubject Signal Pairs. The third strategy is based on the assumption that the neural signals of different subjects are independent. In [28], intersubject EEG signal pair denotes two EEG signals from two different subjects, and intra-subject EEG signal pair denotes two EEG signals from the same subject. Then under this assumption, the intersubject EEG signal pairs can be used as surrogate pairs for intra-subject EEG signal pairs. Compared with above artificial surrogate data, intersubject surrogate pairs seem to possess more inherent features of EEG signals but with no association between each other. Then significance test based on intersubject EEG signal pairs may be used as a standard criterion in evaluating the performance of artificial surrogate methods. Significance test with this strategy shows that the histogram of the MPC-based PSIs λ for intersubject surrogate pairs and intra-subject EEG signal pairs is similar to that based on the

phase-shuffled surrogate method (Figure 2) [28], and the significance threshold suggested by intersubject surrogate pairs is close to that by the phase-shuffled surrogate method. This implies that the phase-shuffled surrogate method is workable in providing significance test for PS analysis. In our opinion, more studies on significance test for PS analysis are needed, and the studies combined the three different strategies introduced above would be promising when they may reach consistent results.

4. Problems in Phase Synchronization Detection

In real applications, the observed neural signals are more or less contaminated by noise, and the samples collected are usually limited. Obviously, the observational noise will degrade the estimation of PSI and even submerge non-trivial PS in neural signals. In addition, the MPC-based PSI is a biased estimator, which implies that the reliability of functional connectivity inferred by it will decrease as the samples collected in signals are insufficient. To get a reliable PS detection in real applications, we must take the issues into consideration. In this section, we would discuss challenges of PS analysis in practice. In particular, we would introduce the advances on the effect of noise in PS detection, the influence of signal duration and estimation bias of PSI, the effect of volume conduction, and the influence of reference in PS analysis.

4.1. Effect of Noise in PS Detection. The effect of noise in PS detection has been examined by both numerical computation [10, 11, 27, 33, 62, 63] and analytical study [32]. A bandpass prefiltering can suppress the effect of noise but may introduce spurious connectivity as well [33]. In this section, we mainly review a theoretical study on the effect of noise in IP estimation and PS detection based on the unified framework of IP definition [32].

Let $s(t) = x(t) + w(t)$ denote the noisy signal, where $x(t)$ is the clean signal and $w(t)$ is the additive noise. The analytic signal of $s(t)$ can be defined with a bandpass filter $b(t)$; that is,

$$\begin{aligned} s^{(b)}(t) &= s(t) * b(t) \\ &= x(t) * b(t) + w(t) * b(t) \\ &= A_x(t) e^{j\phi_x^{(b)}(t)} + w^{(b)}(t), \end{aligned} \quad (21)$$

where $w^{(b)}(t) = w(t) * b(t)$. Let $\theta(t) = \hat{\phi}_s^{(b)}(t) - \phi_x^{(b)}(t)$ denote the estimate error of IP due to the noise term $w(t)$, where $\hat{\phi}_s^{(b)}(t)$ denotes the estimate of $\phi_s^{(b)}(t)$ from the noisy signal $s(t)$. It has been proved that the distribution of $\theta(t)$ is a normal distribution; that is, $\theta \sim N(0, \sigma_\theta^2)$,

$$p(\theta) = \left(\sqrt{2\pi} \sigma_\theta \right)^{-1} e^{-\theta^2/(2\sigma_\theta^2)}, \quad (22)$$

when the instantaneous signal-to-noise ratio, defined as $r^{(b)}(t) = A_x^2(t)/[2\sigma_{w^{(b)}}^2]$ in the pass band, is larger than 5. Here $\sigma_{w^{(b)}}^2$ is the variance of $w^{(b)}(t)$, and $\sigma_\theta = \sigma_{w^{(b)}}/A_x(t)$.

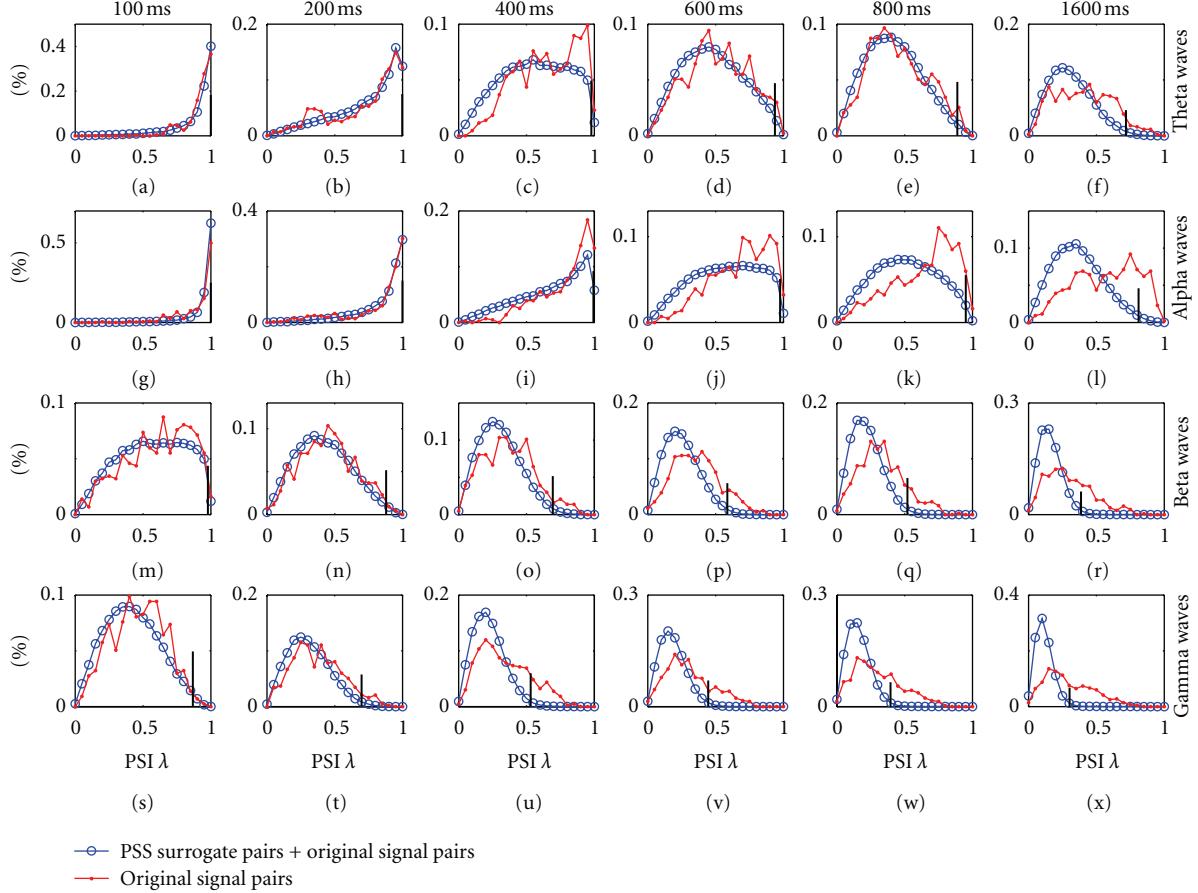


FIGURE 2: Histogram of the phase synchronization indexes (PSI) λ for 435 original EEG signal pairs and their $435 \times 99 = 43\,065$ phase-shuffled surrogate pairs for one subject. The results for the theta, alpha, beta, and gamma waves of six different duration, that is, 100 ms, 200 ms, 400 ms, 600 ms, 800 ms, and 1600 ms, are presented. For each subfigure, all the PSIs λ of original EEG signal pairs and their phase-shuffled surrogate pairs are sorted in ascending order, and the black bar indicates the value (x-axis) of the one at the rank of 95% of all the PSIs. In other words, the black bar indicates the threshold of 95% level of significance for the estimated PSIs of original EEG signal pairs in each case.

The wrapped θ , that is, $\Theta = \theta(\text{mod}2\pi)$, obeys the wrapped normal distribution, $\Theta \sim \tilde{N}(0, \sigma_\theta^2)$ [35, 64],

$$p(\Theta) = \frac{1}{\sqrt{2\pi}\sigma_\theta} \sum_{k=-\infty}^{\infty} e^{-(\Theta+2k\pi)^2/(2\sigma_\theta^2)}. \quad (23)$$

For the MPC-based PSI, $\lambda = |E[e^{j\varphi}]|$, the effect of noise appears as

$$\hat{\lambda} = e^{-(\sigma_{\theta_1}^2 + \sigma_{\theta_2}^2)/2} \lambda; \quad (24)$$

that is, the noise introduces a degrading factor, $e^{-(\sigma_{\theta_1}^2 + \sigma_{\theta_2}^2)/2}$, into the true index λ . This factor is determined by only the level of in-band noise. If $\sigma_\theta = \sigma_{w(b)}/A_x$ in (22) is not a constant, θ actually obeys a conditional distribution, and (22) turns out to be

$$p(\theta | \sigma_\theta) = \left(\sqrt{2\pi}\sigma_\theta \right)^{-1} e^{-\theta^2/(2\sigma_\theta^2)}, \quad \sigma_\theta > 0. \quad (25)$$

For observed signal $\{s(n)\}$, the distribution of IP error is a scale mixture of normal distributions (SMN) with different

variances. If the probability density function of $\{\sigma_\theta(n)\}$ is known, the empirical distribution of IP error $\{\theta(n)\}$ can be approximated as

$$p_m(\theta) = \sum_{k=1}^K p(\theta | \sigma_k) \pi_k, \quad (26)$$

where $\{\pi_k\}_{k=1}^K$ are the respective empirical probabilities which are estimated from $\{A_x(n)\}$ [65]. Note that $A_x(n)$ is the instantaneous amplitude of the clean signal, and thus it is difficult to obtain its distribution with only the observed noisy signal $\{s(n)\}$. Simulations are performed by considering the SMN of the IP error as a normal distribution with constant standard deviation $\sigma_{w(b)}/\max\{A_x(n)\}$; that is, $\sigma_\theta = \sigma_{w(b)}/\max\{A_x(n)\}$, and computational outputs support the theoretical results (24). When the in-band noise level is not so high (signal-to-noise ratio greater than 10 dB), the estimated PSI is not much affected by observational noise [32].

4.2. Influence of Signal Duration and Estimation Bias. Another question in PS analysis is how long a epoch of signals is enough to get a reliable quantification of neural connectivity? One microstate of EEG signals usually lasts 50 to 200 ms [3, 9, 66]. However, various durations (from 100 ms to 10 ms) of EEG signals have been used in examining functional connectivity [12, 24, 67, 68]. A comprehensive study has been performed regarding this issue based on both surrogate tests and intersubject EEG surrogate test [28]. Results show that a duration of EEG waves of 3 ~ 18 wave cycles is suitable for PS analysis. The value of PSI is not only determined by the relationship between two IP sequences but also highly affected by the duration of signals used for PSI estimation. Too short duration of signals will result in too large PSI, while too long duration of signals will yield too small PSI, both of which may not offer a good indication on the level of PS (Figure 2) [28].

In another study, the IP difference of a signal pair with no synchrony is assumed to obey uniform distribution. Under this assumption, the expected value of the MPC-based PSI is

$$E\{\lambda\} \approx \frac{1}{\sqrt{N}}, \quad (27)$$

where N is the number of samples used for PSI estimation [43, 69]. This means that the MPC-based PSI λ is a positively biased estimators for finite sample sizes [70]. More samples used will result in a less biased PSI. Then even for the same duration of signals, signals observed with higher sampling rate will yield a better estimation of the MPC-based PSI than those with lower sampling rate. The effect of sampling rate on both the MPC-based PSI and the entropy-based PSI has been evaluated. Results demonstrate that the MPC-based PSI λ is less affected by sampling rate than the entropy-based PSI and thus is recommended for PS analysis of EEG signals measured with low sampling rate [28].

In many experimental studies, the number of samples or trials of measured data is limited. Bias will be unavoidable in PSI estimation due to insufficient samples. In addition, even when the observed signals are with sufficient samples, they may be nonstationary. In this case, we still want to estimate PSI with samples in short time window, so as to obtain functional connectivity with a certain degree of temporal resolution. Regarding this problem, a new measure called pairwise phase consistency has been proposed [71]. This measure is defined as the mean of the cosine of IP difference across all given signal pairs; that is,

$$\xi = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{k=i+1}^N \cos(\phi_i - \phi_k). \quad (28)$$

Pairwise phase consistency quantifies the similarity of relative IP among trials or samples and is bias-free. In addition, it is demonstrated that pairwise phase consistency is equivalent to the squared MPC-based PSI in population statistic [71].

4.3. Influence of Volume Conduction. The activities of a single source within the brain can be observed by many sensors on the scalp [72]. This is usually referred as volume conduction.

Then the PSI between signals measured by different sensors, especially by spatially adjacent sensors, may be a trivial artefact due to volume conduction, but not a true interaction of the underlying brain activities [73]. There are two ways to tackle this problem. One way is to estimate the true sources of underlying brain activities with observed EEG/MEG signals using inverse method and then quantify the relationship of the estimated sources rather than the relationship of the observed signals. However, the true sources are usually not easy to be obtained by inverse methods.

An alternative way is based on the imaginary part of coherency of observed signals. We assume that two signals $s_i(t)$ and $s_j(t)$ collected by sensors i and j are from a linear superposition of K independent sources $x_k(t)$, and the mapping of each source to sensors is instantaneous with no distortion [74]. Then in the frequency domain, $s_i(t)$ can be expressed as

$$S_i(f) = \sum_{k=1}^K a_{ik} X_k(f), \quad (29)$$

where $S_i(f)$ and $X_k(f)$ are the Fourier transform of $s_i(t)$ and $x_k(t)$, respectively, and a_{ik} are the contribution coefficient of $x_k(t)$ to $s_i(t)$. The cross-spectrum of $s_i(t)$ and $s_j(t)$ is

$$\begin{aligned} C_{ij}(f) &= E\{S_i(f)S_j^*(f)\} \\ &= \sum_{kk'} a_{ik} a_{jk'} E\{X_k(f)X_{k'}^*(f)\} \\ &= \sum_{kk'} a_{ik} a_{jk'} \delta_{kk'} E\{|X_k(f)|^2\} \\ &= \sum_k a_{ik} a_{jk} E\{|X_k(f)|^2\}, \end{aligned} \quad (30)$$

where $\delta_{kk'}$ denotes the Kronecker-delta function. Here $C_{ij}(f)$ is real, which implies that volume conduction of multiple sources strongly affects the real part of the cross-spectrum between $s_i(t)$ and $s_j(t)$ but does not affect the imaginary part [75, 76].

The complex coherency of $s_i(t)$ and $s_j(t)$ is defined as

$$\Omega(f) = \frac{C_{ij}(f)}{\left[C_{ii}(f)C_{jj}^*(f) \right]^{1/2}}. \quad (31)$$

In [75], the imaginary part of $\Omega(f)$ is defined as a synchronization measure for $s_i(t)$ and $s_j(t)$ in the frequency domain. The mean of $\Omega(f)$ over all frequencies is equal to the mean of the cross-correlation of the corresponding analytic signals over time [76]; that is,

$$\langle \Omega(f) \rangle_f = \frac{\left\langle A_i(t)A_j(t)e^{j(\phi_i(t)-\phi_j(t))} \right\rangle_t}{\left[\langle A_i^2(t) \rangle_t \langle A_j^2(t) \rangle_t \right]^{1/2}}, \quad (32)$$

where $\langle \cdot \rangle_f$ and $\langle \cdot \rangle_t$ denote the average, respect to frequency and time respectively. Then the imaginary part of $\langle \Omega(f) \rangle_f$ is

$$\Im[\langle \Omega(f) \rangle_f] = \frac{\left\langle A_i(t)A_j(t) \sin[\phi_i(t) - \phi_j(t)] \right\rangle_t}{\left[\langle A_i^2(t) \rangle_t \langle A_j^2(t) \rangle_t \right]^{1/2}}. \quad (33)$$

However, it has been demonstrated that the imaginary part of $\langle \Omega(f) \rangle_f$ is not a good index of PS as it depends on the amplitudes of signals [76, 77]. Further, a measure called phase lag index,

$$\eta = \left| \left\langle \text{sign}[\phi_i(t) - \phi_j(t)] \right\rangle_t \right|, \quad (34)$$

was proposed based on the consideration that “the existence of a consistent, nonzero phase lag between two times series cannot be explained by volume conduction from a single strong source” [76]. Recently, it is demonstrated that the performance of phase lag index could be degraded by small perturbations [70]. To deal with this problem, a weighted phase lag index is defined as

$$\zeta = \frac{\left| E \left\{ J \left[S_i(f) S_j^*(f) \right] \right\} \right|}{E \left\{ \left| J \left[S_i(f) S_j^*(f) \right] \right| \right\}}. \quad (35)$$

Compared with phase lag index, the weighted phase lag index is demonstrated with “reduced sensitivity to additional, uncorrelated noise sources and increased statistical power to detect changes in phasesynchronization” [70].

4.4. Influence of Reference. The influence of reference on EEG signals is a long-lasting problem in quantifying functional connectivity [78–80]. Various reference strategies, such as bipolar EEG [81], average common reference EEG [82], and Laplacian EEG [73], have been proposed. However, the cautions on these reference strategies have been extensively reported as well [81, 83].

The influence of reference has been examined with both analytical analysis and computational simulation [68]. Let $u(t) = Vr(t)$ denote the reference signal, where $V > 0$ is coefficient and $r(t)$ is time-dependent. Let $y_i(t)$ denote the scalp signal measured by the i th sensor, and, $s_i(t) = u(t) - y_i(t)$ denote the signal $y_i(t)$ rereferenced to $u(t)$. For analytic signal $s_i(t)$ defined with the Hilbert transform, we have

$$\begin{aligned} & \lim_{V \rightarrow +\infty} \phi_{s_i}(t) \\ &= \lim_{V \rightarrow +\infty} \arctan \frac{\tilde{s}_i(t)}{s_i(t)} \\ &= \lim_{V \rightarrow +\infty} \arctan \frac{(1/\pi)P.V. \int_{-\infty}^{+\infty} [(Vr(\tau)/t - \tau) - (y_i(\tau)/t - \tau)] d\tau}{Vr(t) - y_i(t)} \\ &= \lim_{V \rightarrow +\infty} \arctan \frac{(1/\pi)P.V. \int_{-\infty}^{+\infty} [(Vr(\tau)/t - \tau) - (y_i(\tau)/t - \tau)] d\tau}{Vr(t)} \\ &= \lim_{V \rightarrow +\infty} \arctan \frac{(1/\pi)P.V. \int_{-\infty}^{+\infty} (Vr(\tau)/t - \tau) d\tau}{Vr(t)} \\ &= \lim_{V \rightarrow +\infty} \arctan \frac{(1/\pi)P.V. \int_{-\infty}^{+\infty} (r(\tau)/t - \tau) d\tau}{r(t)}. \end{aligned} \quad (36)$$

Then we have

$$\lim_{V \rightarrow +\infty} \lambda_{s_1 s_2} = \lim_{V \rightarrow +\infty} \left| E \left\{ e^{j[\phi_{s_1}(t) - \phi_{s_2}(t)]} \right\} \right| = 1, \quad (37)$$

as $\lim_{V \rightarrow +\infty} \phi_{s_1}(t) - \lim_{V \rightarrow +\infty} \phi_{s_2}(t) = 0$. This result indicates that the coefficient V has a great influence on the MPC-based PSI λ . A sufficiently large V will lead to larger PSI

$\lambda_{s_1 s_2}$ for referential signals than that for nonreferential signals [68]. Simulation study demonstrates that the PSI $\lambda_{s_1 s_2}$ of two referential signals may monotonically increase to 1 or decrease first and then increase to 1 as the coefficient V increases from 0 to ∞ . In addition, a method based on independent component analysis has been demonstrated to be an appropriate method to generate reference for EEG signals in quantifying connectivity [84].

5. Extensions of Phase Synchronization Analysis

Various extensions of PS analysis have been proposed to infer the relationship in multivariate or multitrial signals based on the concept of IP [85–90]. For example, a method named frequency flow analysis has been used to examine the global synchronization of multivariate signals [88]. If the IF, derived from IP, of each variable almost equals to the IFs of other variables in a certain frequency band for a duration, the multivariate signals are called in PS accordingly. In this section, we will further introduce two extensions/variations of PS analysis, that is, the trial-based PSI [85, 86] and the partial PSI [87].

A trial-based PSI is proposed to examine the variation of IP difference between signal channels across trials under repeated stimulus [85, 86]. For a data set of K trials and each trial with N samples, the trial-based PSI is defined as

$$\lambda_{s_i s_j}(t) = \frac{1}{K} \sum_{k=1}^K e^{j[\phi_{s_i}(t, k) - \phi_{s_j}(t, k)]}, \quad (38)$$

where $\phi_{s_i}(t, k)$ and $\phi_{s_j}(t, k)$ are the IPs of the channels s_i and s_j in the k th trial. The PSI $\lambda_{s_i s_j}(t)$ measures the intertrial variability of IP difference between channels s_1 and s_2 at instant t . The PSI $\lambda_{s_i s_j}(t)$ with value close to 1 implies that the IP difference varies little across trials at instant t , while the PSI $\lambda_{s_i s_j}(t)$ with value close to zero means that the IP difference varies approximately uniformly across trials.

The pairwise PSI discussed above can infer the strength of connectivity in two signals but cannot indicate whether the connectivity is induced by the direct coupling between them or due to the indirect interaction mediated by other units. To deal with this problem, a measure called partial PSI is generalized from PS analysis following the idea of partial coherence [87, 91]. Combined with the pairwise PSI, the partial PSI can be used to distinguish the direct and indirect interdependencies among interacted systems/units [87]. For a set of time series $\{s_i(n)\}$, $i = 1, 2, \dots, L$, a matrix is defined with PSIs of bivariate signals as

$$\mathbf{R} = \begin{pmatrix} 1 & \lambda_{s_1 s_2} & \dots & \lambda_{s_1 s_L} \\ \lambda_{s_1 s_2}^* & 1 & \dots & \lambda_{s_2 s_L} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{s_1 s_L}^* & \lambda_{s_2 s_L}^* & \dots & 1 \end{pmatrix}. \quad (39)$$

Let $\mathbf{\Gamma} = \mathbf{R}^{-1}$ denote the inverse matrix of \mathbf{R} . Then a measure called partial PSI is defined as

$$\lambda_{ij|Z} = \frac{|\mathbf{\Gamma}_{ij}|}{[\mathbf{\Gamma}_{ii}\mathbf{\Gamma}_{jj}]^{1/2}}, \quad (40)$$

for $\{s_i(n)\}$ and $\{s_j(n)\}$, conditioning on the remaining processes $\{S_Z \mid Z = 1, 2, \dots, L, Z \neq i, j\}$ [87]. A partial PSI $\lambda_{ij|Z} \approx 0$ would imply that the association is induced by indirect coupling if the pair-wise PSI $\lambda_{s_i s_j}$ between $\{s_i(n)\}$ and $\{s_j(n)\}$ is significant.

6. Discussions and Conclusions

We give a technical review on PS analysis in this paper. In particular, we discuss IP definitions, PSI estimation and its significance test, the issues that may affect PS detection, and extensions of PSI. PS analysis is a method to quantify the mutual rhythmic interaction of coupled systems/units and has been used to infer functional connectivity from observed neural signals such as EEG and MEG. The main advantage of PS analysis is that it could detect weak interaction between signal pairs by only taking the IPs of signals into consideration but neglecting the influence of instantaneous amplitudes of signals. In addition, PS analysis could work for nonstationary signals. These merits imply that PS analysis is suitable for neuroscience research, as we are usually interested in the relationship between neural oscillations in particular frequency bands such as beta waves ([12, 30] Hz) and gamma waves ([30, 80] Hz) rather than the interaction between broadband raw signals.

While inferring functional neural connectivity with PS analysis, several cautions and limitations should be taken into consideration. First, the observed neural signal is usually with broadband spectra and unavoidably contaminated by noise. In this case, bandpass filter should be used to extract neural oscillations in the raw signals. Second, for spatially adjacent neural recordings, PS would be affected by volume conduction and so did correlation coefficient and mutual information. In this case, the phase lag index [76] or the weighted phase lag index [70] is recommended. Third, PS analysis does not work in a black-box way. The PSIs estimated with different durations of neural signals are not recommended to be compared. Fourth, users should be aware that PSI quantifies the variation instead of the mean of IP difference within a period. Therefore, PS analysis might not be suitable for analyzing those relatively stable components in neural signals, such as event-related potentials estimated from multiple EEG trials, as PS would ignore the amplitudes of components in event-related potential and miss the latencies of these components as well.

Various measures, such as cross-correlation, coherence, nonlinear interdependence [10], mutual information [8], partial directed coherence [9], correlation-entropy coefficient [92], and coherence entropy coefficient [13], have been proposed in functional connectivity analysis based on EEG, MEG signals and simulated data from various aspects [10–13, 93–95]. Results show that they can reveal a similar

tendency of global connectivity [10]. The MPC-based PSI has been argued to be slightly better than coherence for both estimation and detection purposes [93]. In another study, total 34 different measures are classified into several families such as correlation/coherence family, mutual information family, PS family, and the Granger causality family. These measures are comprehensively compared, and results show that PS measures, Granger's causality measures and stochastic event synchrony measures are only weakly correlated with correlation coefficient, and are mutually uncorrelated as well [13]. These results imply that these measures each could characterize interdependence of signals from different aspects. With this consideration, we could have a good characterization of functional connectivity between neural signals with only a few representative measures of different families.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 60901025, 61174045, and 60935001), National Basic Research Program of China (973 Program) (no. 2011CB013304), and the International Science and Technology Cooperation Program of China (no. 2011DFA10950).

References

- [1] G. Tononi, O. Sporns, and G. M. Edelman, "A measure for brain complexity: relating functional segregation and integration in the nervous system," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 11, pp. 5033–5037, 1994.
- [2] P. Tass, M. G. Rosenblum, J. Weule et al., "Detection of $n:m$ phase locking from noisy data: application to magnetoencephalography," *Physical Review Letters*, vol. 81, no. 15, pp. 3291–3294, 1998.
- [3] F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: phase synchronization and large-scale integration," *Nature Reviews Neuroscience*, vol. 2, no. 4, pp. 229–239, 2001.
- [4] R. T. Canolty, E. Edwards, S. S. Dalal et al., "High gamma power is phase-locked to theta oscillations in human neocortex," *Science*, vol. 313, no. 5793, pp. 1626–1628, 2006.
- [5] J. Fell and N. Axmacher, "The role of phase synchronization in memory processes," *Nature Reviews Neuroscience*, vol. 12, no. 2, pp. 105–118, 2011.
- [6] B. He, L. Yang, C. Wilke, and H. Yuan, "Electrophysiological imaging of brain activity and connectivity—challenges and opportunities," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 7, pp. 1918–1931, 2011.
- [7] P. J. Uhlhaas and W. Singer, "Abnormal neural oscillations and synchrony in schizophrenia," *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 100–113, 2010.
- [8] J. M. Hurtado, L. L. Rubchinsky, and K. A. Sigvardt, "Statistical method for detection of phase-locking episodes in neural oscillations," *Journal of Neurophysiology*, vol. 91, no. 4, pp. 1883–1898, 2004.
- [9] M. Ding, S. L. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials

- by adaptive multivariate autoregressive modeling: data pre-processing, model validation, and variability assessment," *Biological Cybernetics*, vol. 83, no. 1, pp. 35–45, 2000.
- [10] R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, "Performance of different synchronization measures in real data: a case study on electroencephalographic signals," *Physical Review E*, vol. 65, no. 4, Article ID 041903, 14 pages, 2002.
- [11] T. Kreuz, F. Mormann, R. G. Andrzejak, A. Kraskov, K. Lehnertz, and P. Grassberger, "Measuring synchronization in coupled model systems: a comparison of different approaches," *Physica D*, vol. 225, no. 1, pp. 29–42, 2007.
- [12] L. Astolfi, F. De Vico Fallani, F. Cincotti et al., "Estimation of effective and functional cortical connectivity from neuroelectric and hemodynamic recordings," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 3, pp. 224–233, 2009.
- [13] J. Dauwels, F. Vialatte, T. Musha, and A. Cichocki, "A comparative study of synchrony measures for the early diagnosis of alzheimer's disease based on EEG," *Neuroimage*, vol. 49, no. 1, pp. 668–693, 2010.
- [14] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge University Press, Cambridge, UK, 2001.
- [15] R. Brown and L. Kocarev, "A unifying definition of synchronization for dynamical systems," *Chaos*, vol. 10, no. 2, pp. 344–349, 2000.
- [16] M. Le Van Quyen and A. Bragin, "Analysis of dynamic brain oscillations: methodological advances," *Trends in Neurosciences*, vol. 30, no. 7, pp. 365–373, 2007.
- [17] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Physical Review Letters*, vol. 76, no. 11, pp. 1804–1807, 1996.
- [18] D. J. DeShazer, R. Breban, E. Ott, and R. Roy, "Detecting phase synchronization in a chaotic laser array," *Physical Review Letters*, vol. 87, no. 4, Article ID 044101, 4 pages, 2001.
- [19] I. Z. Kiss, Q. Lv, and J. L. Hudson, "Synchronization of non-phase-coherent chaotic electrochemical oscillations," *Physical Review E*, vol. 71, no. 3, Article ID 035201, 4 pages, 2005.
- [20] M. G. Kitzbichler, M. L. Smith, S. R. Christensen, and E. Bullmore, "Broadband criticality of human brain network synchronization," *Plos Computational Biology*, vol. 5, no. 3, article e1000314, 2009.
- [21] M. Chávez, M. Le Van Quyen, V. Navarro, M. Baulac, and J. Martinerie, "Spatio-temporal dynamics prior to neocortical seizures: amplitude versus phase couplings," *IEEE Transactions on Bio-medical Engineering*, vol. 50, no. 5, pp. 571–583, 2003.
- [22] W. Wu, J. Sun, Z. Jin et al., "Impaired neuronal synchrony after focal ischemic stroke in elderly patients," *Clinical Neurophysiology*, vol. 122, no. 1, pp. 21–26, 2011.
- [23] E. Gysels and P. Celka, "Phase synchronization for the recognition of mental tasks in a brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 4, pp. 406–415, 2004.
- [24] C. Brunner, R. Scherer, B. Graimann, G. Supp, and G. Pfurtscheller, "Online control of a brain-computer interface using phase synchronization," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2501–2506, 2006.
- [25] P. J. Uhlhaas and W. Singer, "Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology," *Neuron*, vol. 52, no. 1, pp. 155–168, 2006.
- [26] C. Hammond, H. Bergman, and P. Brown, "Pathological synchronization in parkinson's disease: networks, models and treatments," *Trends in Neurosciences*, vol. 30, no. 7, pp. 357–364, 2007.
- [27] M. Winterhalder, B. Schelter, J. Kurths, A. Schulze-Bonhage, and J. Timmer, "Sensitivity and specificity of coherence and phase synchronization analysis," *Physics Letters Section A*, vol. 356, no. 1, pp. 26–34, 2006.
- [28] J. Sun, X. Hong, and S. Tong, "Phase synchronization analysis of EEG signals: an evaluation based on surrogate tests," *IEEE Transactions on Biomedical Engineering*. Submitted.
- [29] R. A. Dobie and M. J. Wilson, "Analysis of auditory evoked potentials by magnitude-squared coherence," *Ear and Hearing*, vol. 10, no. 1, pp. 2–13, 1989.
- [30] S. Kalitzin, J. Parra, D. N. Velis, and F. H. Lopes Da Silva, "Enhancement of phase clustering in the EEG/meg gamma frequency band anticipates transitions to paroxysmal epileptiform activity in epileptic patients with known visual sensitivity," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 11, pp. 1279–1286, 2002.
- [31] A. E. Hramov and A. A. Koronovskii, "An approach to chaotic synchronization," *Chaos*, vol. 14, no. 3, pp. 603–610, 2004.
- [32] J. Sun and M. Small, "Unified framework for detecting phase synchronization in coupled time series," *Physical Review E*, vol. 80, no. 4, Article ID 046219, 2009.
- [33] L. Xu, Z. Chen, K. Hu, H. E. Stanley, and P. C. Ivanov, "Spurious detection of phase synchronization in coupled nonlinear oscillators," *Physical Review E*, vol. 73, no. 6, Article ID 065201, 2006.
- [34] B. Schelter, M. Winterhalder, J. Timmer, and M. Peifer, "Testing for phase synchronization," *Physics Letters Section A*, vol. 366, no. 4-5, pp. 382–390, 2007.
- [35] S. R. Jammalamadaka and A. SenGupta, *Topics in Circular Statistics*, World Scientific Publishing, Singapore, 2001.
- [36] E. Bedrosian, "A product theorem for Hilbert transform," *Proceedings of the IEEE*, vol. 51, no. 5, pp. 868–869, 1963.
- [37] P. J. Loughlin, "Spectrographic measurement of instantaneous frequency and the time-dependent weighted average instantaneous frequency," *Journal of the Acoustical Society of America*, vol. 105, no. 1, pp. 264–274, 1999.
- [38] P. M. Oliveira and V. Barroso, "Definitions of instantaneous frequency under physical constraints," *Journal of the Franklin Institute*, vol. 337, no. 4, pp. 303–316, 2000.
- [39] T. Pereira, M. S. Baptista, and J. Kurths, "Phase and average period of chaotic oscillators," *Physics Letters Section A*, vol. 362, no. 2–3, pp. 159–165, 2007.
- [40] G. V. Osipov, B. Hu, C. Zhou, M. V. Ivanchenko, and J. Kurths, "Three types of transitions to phase synchronization in coupled chaotic oscillators," *Physical Review Letters*, vol. 91, no. 2, Article ID 024101, 4 pages, 2003.
- [41] D. Vakman, "Computer measuring of frequency stability and the analytic signal," *IEEE Transactions on Instrumentation and Measurement*, vol. 43, no. 4, pp. 668–671, 1994.
- [42] A. H. Nuttall and E. Bedrosian, "On the quadrature approximation to the hilbert transform of modulated signals," *Proceedings of the IEEE*, vol. 54, no. 10, pp. 1458–1459, 1966.
- [43] S. Aviyente and A. Y. Mutlu, "A time-frequency-based approach to phase and phase synchrony estimation," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3086–3098, 2011.
- [44] M. Wacker and H. Witte, "Adaptive phase extraction: incorporating the Gabor transform in the matching pursuit algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2844–2851, 2011.
- [45] D. Vakman, "On the analytic signal the teager-kaiser energy algorithm and other methods for defining amplitude and frequency," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 791–797, 1996.

- [46] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.
- [47] D. Hoyer, O. Hoyer, and U. Zwiener, "A new approach to uncover dynamic phase coordination and synchronization," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 1, pp. 68–74, 2000.
- [48] M. Wacker and H. Witte, "On the stability of the n:m phase synchronization index," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 332–338, 2011.
- [49] M. Paluš, "Detecting phase synchronization in noisy systems," *Physics Letters Section A*, vol. 235, no. 4, pp. 341–351, 1997.
- [50] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [51] E. T. Bullmore and D. S. Bassett, "Brain graphs: graphical models of the human brain connectome," *Annual Review of Clinical Psychology*, vol. 7, pp. 113–140, 2011.
- [52] D. S. Bassett, E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger, and A. Meyer-Lindenberg, "Hierarchical organization of human cortical networks in health and schizophrenia," *Journal of Neuroscience*, vol. 28, no. 37, pp. 9239–9248, 2008.
- [53] Y. Liu, M. Liang, Y. Zhou et al., "Disrupted small-world networks in schizophrenia," *Brain*, vol. 131, no. 4, pp. 945–961, 2008.
- [54] K. Wu, Y. Taki, K. Sato et al., "Age-related changes in topological organization of structural brain networks in healthy individuals," *Human Brain Mapping*, vol. 33, no. 3, pp. 552–568, 2012.
- [55] J. Sun, S. Tong, and G.-Y. Yang, "Reorganization of brain networks in aging and age-related diseases," *Aging and Disease*. In press.
- [56] C. Alfeld and J. Kurths, "Testing for phase synchronization," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 14, no. 2, pp. 405–416, 2004.
- [57] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, no. 3–4, pp. 346–382, 2000.
- [58] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D*, vol. 58, no. 1–4, pp. 77–94, 1992.
- [59] R. G. Andrzejak, A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz, "Bivariate surrogate techniques: necessity, strengths, and caveats," *Physical Review E*, vol. 68, no. 6, Article ID 066202, 5 pages, 2003.
- [60] L. Faes, G. D. Pinna, A. Porta, R. Maestri, and G. Nollo, "Surrogate data analysis for assessing the significance of the coherence function," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1156–1166, 2004.
- [61] M. Thiel, M. C. Romano, J. Kurths, M. Rolfs, and R. Kliegl, "Twin surrogates to test for complex synchronisation," *Europhysics Letters*, vol. 75, no. 4, pp. 535–541, 2006.
- [62] J. Kurths, M. C. Romano, M. Thiel et al., "Synchronization analysis of coupled noncoherent oscillators," *Nonlinear Dynamics*, vol. 44, no. 1–4, pp. 135–149, 2006.
- [63] J. Sun, J. Zhang, J. Zhou, X. Xu, and M. Small, "Detecting phase synchronization in noisy data from coupled chaotic oscillators," *Physical Review E*, vol. 77, no. 4, Article ID 046213, 2008.
- [64] B. C. Lovell and R. C. Williamson, "The statistical performance of some instantaneous frequency estimators," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1708–1723, 1992.
- [65] H. Hamdan, "Characterizing and approximating infinite scale mixtures of normals," *Communications in Statistics*, vol. 35, no. 3, pp. 407–413, 2006.
- [66] D. Van De Ville, J. Britz, and C. M. Michel, "EEG microstate sequences in healthy humans at rest reveal scale-free dynamics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 42, pp. 18179–18184, 2010.
- [67] P. J. Uhlhaas, D. E. J. Linden, W. Singer et al., "Dysfunctional long-range coordination of neural activity during gestalt perception in schizophrenia," *Journal of Neuroscience*, vol. 26, no. 31, pp. 8168–8175, 2006.
- [68] S. Hu, M. Stead, Q. Dai, and G. A. Worrell, "On the recording reference contribution to EEG correlation, phase synchrony, and coherence," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 40, no. 5, pp. 1294–1304, 2010.
- [69] G. V. Tcheslavski, *Coherence and phase synchrony analysis of electroencephalogram*, Ph.D. dissertation, the Virginia Polytechnic Institute, State University, Blacksburg, Va, USA, 2005.
- [70] M. Vinck, R. Oostenveld, M. Van Wingerden, F. Battaglia, and C. M. A. Pennartz, "An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias," *Neuroimage*, vol. 55, no. 4, pp. 1548–1565, 2011.
- [71] M. Vinck, M. van Wingerden, T. Womelsdorf, P. Fries, and C. M. A. Pennartz, "The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization," *Neuroimage*, vol. 51, no. 1, pp. 112–122, 2010.
- [72] J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Physics in Medicine and Biology*, vol. 32, no. 1, article 004, pp. 11–22, 1987.
- [73] P. L. Nunez, R. Srinivasan, A. F. Westdorp et al., "EEG coherency i: statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 5, pp. 499–515, 1997.
- [74] J. G. Stinstra and M. J. Peters, "The volume conductor may act as a temporal filter on the ecg and EEG," *Medical and Biological Engineering and Computing*, vol. 36, no. 6, pp. 711–716, 1998.
- [75] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.
- [76] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: assessment of functional connectivity from multi channel EEG and meg with diminished bias from common sources," *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007.
- [77] L. A. Wheaton, G. Nolte, S. Bohlhalter, E. Fridman, and M. Hallett, "Synchronization of parietal and premotor areas during preparation and execution of praxis hand movements," *Clinical Neurophysiology*, vol. 116, no. 6, pp. 1382–1390, 2005.
- [78] L. T. Trujillo, M. A. Peterson, A. W. Kaszniak, and J. J. B. Allen, "EEG phase synchrony differences across visual perception conditions may depend on recording and analysis methods," *Clinical Neurophysiology*, vol. 116, no. 1, pp. 172–189, 2005.
- [79] S. J. Schiff, "Dangerous phase," *Neuroinformatics*, vol. 3, no. 4, pp. 315–317, 2005.
- [80] R. Guevara, J. L. P. Velazquez, V. Nenadovic, R. Wennberg, G. Senjanović, and L. G. Dominguez, "Phase synchronization measurements using electroencephalographic recordings: what can we really say about neuronal synchrony?" *Neuroinformatics*, vol. 3, no. 4, pp. 301–313, 2005.
- [81] H. P. Zaveri, R. B. Duckrow, and S. S. Spencer, "On the use of bipolar montages for time-series analysis of intracranial

- electroencephalograms,” *Clinical Neurophysiology*, vol. 117, no. 9, pp. 2102–2108, 2006.
- [82] O. Bertrand, F. Perrin, and J. Pernier, “A theoretical justification of the average reference in topographic evoked potential studies,” *Electroencephalography and Clinical Neurophysiology*, vol. 62, no. 6, pp. 462–464, 1985.
- [83] H. P. Zaveri, R. B. Duckrow, and S. S. Spencer, “The effect of a scalp reference signal on coherence measurements of intracranial electroencephalograms,” *Clinical Neurophysiology*, vol. 111, no. 7, pp. 1293–1299, 2000.
- [84] S. Hu, M. Stead, and G. A. Worrell, “Automatic identification and removal of scalp reference signal for intracranial EEGs based on independent component analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 9, pp. 1560–1572, 2007.
- [85] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, “Measuring phase synchrony in brain signals,” *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.
- [86] L. Li, J. X. Zhang, and T. Jiang, “Visual working memory load-related changes in neural activity and functional connectivity,” *Plos One*, vol. 6, no. 7, Article ID e22357, 2011.
- [87] B. Schelter, M. Winterhalder, R. Dahlhaus, J. Kurths, and J. Timmer, “Partial phase synchronization for multivariate synchronizing systems,” *Physical Review Letters*, vol. 96, no. 20, Article ID 208103, 2006.
- [88] D. Rudrauf, A. Douiri, C. Kovach et al., “Frequency flows and the time-frequency dynamics of multivariate phase synchronization in brain signals,” *Neuroimage*, vol. 31, no. 1, pp. 209–227, 2006.
- [89] H. Osterhage, F. Mormann, T. Wagner, and K. Lehnertz, “Detecting directional coupling in the human epileptic brain: limitations and potential pitfalls,” *Physical Review E*, vol. 77, no. 1, Article ID 011914, 2008.
- [90] C. Allefeld and S. Bialonski, “Detecting synchronization clusters in multivariate time series via coarse-graining of markov chains,” *Physical Review E*, vol. 76, no. 6, Article ID 066207, 2007.
- [91] R. Dahlhaus, “Graphical interaction models for multivariate time series,” *Metrika*, vol. 51, no. 2, pp. 157–172, 2000.
- [92] A. Gunduz and J. C. Principe, “Correntropy as a novel measure for nonlinearity tests,” *Signal Processing*, vol. 89, no. 1, pp. 14–23, 2009.
- [93] A. V. Sazonov, C. K. Ho, J. W. M. Bergmans et al., “An investigation of the phase locking index for measuring of interdependency of cortical source signals recorded in the EEG,” *Biological Cybernetics*, vol. 100, no. 2, pp. 129–146, 2009.
- [94] A. Bruns, “Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches?” *Journal of Neuroscience Methods*, vol. 137, no. 2, pp. 321–332, 2004.
- [95] O. David, D. Cosmelli, and K. J. Friston, “Evaluation of different measures of functional connectivity using a neural mass model,” *Neuroimage*, vol. 21, no. 2, pp. 659–673, 2004.