ARTICLE IN PRESS

NeuroImage xxx (2017) 1–13

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Hyperedge bundling: A practical solution to spurious interactions in MEG/EEG source connectivity analyses

Sheng H. Wang [a,b,c,*], Muriel Lobier [a], Felix Siebenhühner [a,b], Tuomas Puoliväli [a,b], Satu Palva [a,c], J. Matias Palva [a,**]

[a] Neuroscience Center, Helsinki Institute of Life Science (HiLife), University of Helsinki, Finland
[b] Doctoral Programme Brain & Mind, University of Helsinki, Finland
[c] BioMag Laboratory, HUS Medical Imaging Center, Helsinki, Finland

ARTICLE INFO

Keywords:
Signal leakage
Spurious correlation
Artificial correlation
Volume conduction
Signal mixing
Point spread
Graph theory
MEG
EEG

ABSTRACT

Inter-areal functional connectivity (FC), neuronal synchronization in particular, is thought to constitute a key systems-level mechanism for coordination of neuronal processing and communication between brain regions. Evidence to support this hypothesis has been gained largely using invasive electrophysiological approaches. In humans, neuronal activity can be non-invasively recorded only with magneto- and electroencephalography (MEG/EEG), which have been used to assess FC networks with high temporal resolution and whole-scalp coverage. However, even in source-reconstructed MEG/EEG data, signal mixing, or "source leakage", is a significant confounder for FC analyses and network localization.

Signal mixing leads to two distinct kinds of false-positive observations: artificial interactions (AI) caused directly by mixing and spurious interactions (SI) arising indirectly from the spread of signals from true interacting sources to nearby false loci. To date, several interaction metrics have been developed to solve the AI problem, but the SI problem has remained largely intractable in MEG/EEG all-to-all source connectivity studies. Here, we advance a novel approach for correcting SIs in FC analyses using source-reconstructed MEG/EEG data.

Our approach is to bundle observed FC connections into hyperedges by their adjacency in signal mixing. Using realistic simulations, we show here that bundling yields hyperedges with good separability of true positives and little loss in the true positive rate. Hyperedge bundling thus significantly decreases graph noise by minimizing the false-positive to true-positive ratio. Finally, we demonstrate the advantage of edge bundling in the visualization of large-scale cortical networks with real MEG data. We propose that hypergraphs yielded by bundling represent well the set of true cortical interactions that are detectable and dissociable in MEG/EEG connectivity analysis.

Introduction

Large-scale neuronal networks, *e.g.*, manifested by functional, directed, and effective connectivity (Karl, 2011), are thought to be critical for healthy brain functions while their abnormalities are thought to underlie many brain diseases (Brookes et al., 2016; Bullmore and Sporns, 2009, 2012; Fornito et al., 2015; Papo et al., 2014; Petersen and Sporns, 2015; Rubinov 2015; Sporns, 2014; Uhlhaas and Singer 2006, 2010). Currently, magneto- and electro-encephalography (MEG/EEG) are the only non-invasive electrophysiological tools for studying connectivity networks with millisecond-range temporal resolution and good coverage of the cortical surface (Kujala et al., 2008; Palva and Palva, 2012; S. Baillet et al., 2001; Salmelin and Baillet, 2009). Accurately identifying interaction dynamics from MEG/EEG data is of crucial importance for understanding their role in human cognition and its deficits.

To date, numerous interaction metrics have been developed and utilized to assess functional connectivity (FC) in terms of amplitude-, phase-, and phase-amplitude correlations within or across frequency bands for pairs of electrophysiological signals (Bastos and Schoffelen, 2016; Kreuz, 2011; O'Neill et al., 2015). These pairwise metrics are typically applied to estimate FC among all brain regions, *i.e.*, to obtain "all-to-all" FC connectomes (Sporns et al., 2005). Networks of inter-areal FC are often represented as graphs where brain areas constitute the *nodes*

* Corresponding author. Neuroscience Center, University of Helsinki, Finland.
** Corresponding author.
E-mail addresses: sheng.wang@helsinki.fi (S.H. Wang), matias.palva@helsinki.fi (J.M. Palva).

https://doi.org/10.1016/j.neuroimage.2018.01.056
Received 18 November 2017; Received in revised form 19 January 2018; Accepted 21 January 2018
Available online xxxx
1053-8119/

Please cite this article in press as: Wang, S.H., et al., Hyperedge bundling: A practical solution to spurious interactions in MEG/EEG source connectivity analyses, NeuroImage (2017), https://doi.org/10.1016/j.neuroimage.2018.01.056

(or vertices) and observed inter-areal connections the *edges* (Bullmore and Sporns, 2009; Rubinov and Sporns, 2010).

FC graphs estimated from MEG/EEG sensor space data are neuro-anatomically uninformative and severely confounded by signal mixing. Signal mixing has two facets: first, any focal neuronal signal is picked up by several sensors. Conversely, one sensor detects a mixture of signals from several distinct sources. Source reconstruction can be used to reduce signal mixing and, importantly, elucidate the likely neuroanatomical sources of the MEG/EEG signals (Buzsaki et al., 2012; Gross et al., 2013; Hamalainen et al., 1993; Palva and Palva, 2012; Schoffelen and Gross, 2009). Yet, because of ill-posed nature of the inverse problem, no source reconstruction approach can yield an unambiguous estimate of the source topography. Residual signal mixing in source space, signal leakage, is quantitatively dependent on the source-reconstruction method of choice but qualitatively characteristic to all such methods.

Because of signal leakage, FC measures exhibit two distinct types of false positive observations: *artificial interactions* (AI) and *spurious interactions* (SI), see Box 2 in Palva and Palva (2012). AIs arise directly from the signal mixing by one true signal being smeared to multiple sensors or sources, regardless of whether true interactions are present. SIs are "ghost" interactions caused by the leakage of the signals from two true connected nodes to their surroundings nodes that in turn become falsely connected like the truly connected nodes (Colclough et al., 2015; Farahibozorg et al., 2017; Korhonen et al., 2014; Palva et al., 2017; Palva and Palva, 2012). AIs can be suppressed by a number of bivariate metrics that typically aim to remove linear coupling terms, and therefore removing artificial and true interactions with zero- and anti-phase-lag coupling (for a review see (Palva et al., 2017)). However, the problem of SIs is much less acknowledged and more difficult to solve because SIs stem from multivariate mixing effects. With typical distributed source modeling approaches, signal leakage causes a large number of SIs that render both the network localization and graph property estimates inaccurate (Drakesmith et al., 2015). To date, one solution has been proposed for correcting SIs in oscillation amplitude correlation estimates, which simultaneously orthogonalizes all source time series through the Löwdin procedure (Colclough et al., 2015, 2016). Despite this promising advance, no solutions have yet been proposed to suppress SIs for other interaction metrics.

Here we advance a novel approach, hyperedge bundling, to alleviate the SIs problem in connectivity analyses performed with any interaction metric. Instead of correcting the mixing effects in source signals *per se*, the approach is based on quantifying the extent of mixing between all sources, evaluation of mixing similarity among all edges, and then clustering the *raw* interaction metric edges into *hyperedge* bundles. This procedure aims to yield a hypergraph where each hyperedge represents a true interaction and its spurious reflections.

In this study, we performed a large set of connectivity simulations and realistic all-to-all MEG source space analyses, in which we estimated phase synchrony as a measure of FC with an AI-insensitive metric. We show that in simulated graphs, hyperedge bundling greatly decreases the number of false positives, *i.e.*, SIs. We illustrated how bundling can support an informative visualization of FC graphs with real MEG data. We suggest that such hypergraphs constitute accurate and unbiased representations of neuronal interactions observable in MEG/EEG source space.

Theory

This section covers general topics as follows: signal mixing in MEG/EEG, how spurious interactions (SI) arise from mixing between sources; and bundling of raw edges into hyperedges. The implementations specific to this study are described in the *Methods* section. Throughout the report, we denote a connectivity graph estimated from reconstructed source time series as raw graph $G_{raw} = (V, E)$, where brain regions are nodes $v_i \in V$ and interactions between nodes are "raw" edges, $e_k = \{(v_i, v_j) \in E | v_i, v_i \in V\}$.

Signal mixing results in false positive artificial (AI) and spurious interactions (SI)

Let us consider a scenario where a true phase correlation is present between two distant (unmixed) sources V_1 and V_2 (Fig. 1A top). The signals from V_1 and V_2 are mixed with signals of their nearby and mutually uncorrelated neighbors V_3 and V_4. Estimating phase FC among all four nodes with the phase-locking value (*PLV*) will reveal both the true edge $E(V_1,V_2)$ and false positive "short-range" AIs between the nearby nodes $E(V_1,V_3)$ and $E(V_2,V_4)$, because *PLV* is inflated by mixing (thick gray edges, Fig. 1A bottom). However, due to leakage of the signal from V_1 and V_2 to their neighbors V_3 and V_4, false positive "long-range" SIs $E(V_3,V_4)$, $E(V_2,V_3)$, and $E(V_1,V_4)$ will also be observed (thin dashed edges). These SIs are thus only indirectly caused by mixing and, unlike the zero-phase-lag AIs (see 2.2), SIs inherit the phase-lag of the true interaction (Colclough et al., 2015; Farahibozorg et al., 2017). Mixing-insensitive bivariate metrics such as the imaginary part of *PLV* (*iPLV*) can remove AIs but do not eliminate SIs if the true coupling has non-zero phase lag.

Quantifying the mixing between reconstructed sources

Signal mixing/leakage between two sources is instantaneous and therefore always leads to inflated zero-phase-lag correlations between the sources. Mixing does not vary over time or across frequency bands (Brookes et al., 2012, 2014; Drakesmith et al., 2013; Nolte et al., 2004; Palva and Palva, 2012).

Source-reconstruction

Suppose we have a data matrix $X = \{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\} \in \mathbb{R}^{n\times t}$ representing narrow-band time series of t samples from n neuronal populations. Simulating a MEG/EEG recording, X can be linearly projected to sensor-space (Hämäläinen and Ilmoniemi, 1994):

$$Y = \Gamma X + \varepsilon \tag{1}$$

where $Y \in \mathbb{R}^{s\times t}$ represents the forward-modeled time series from s sensors ($n > s$). Here, $\Gamma \in \mathbb{R}^{s\times n}$ is the forward operator (or the lead field) and $\varepsilon \in \mathbb{R}^{s\times t}$ is the model prediction error derived from measurement noise. Next, Y can be projected back into the source-space, *e.g.*, by minimum-norm estimation (MNE) based inverse modeling:

$$\hat{X} = WY = R\Gamma^T\left(\Gamma R \Gamma^T + \lambda^2\chi\right)^{-1} Y \tag{2}$$

where $W \in \mathbb{R}^{n\times s}$ is the inverse operator (sources × sensors), the regularization parameter $\lambda^{2=}0.1$, R is the source covariance matrix, and χ is the noise covariance matrix. After inverse modeling, the 5000–10000 source time series are collapsed into parcel time series for a cortical parcellation with 50–400 parcels. In the present study, we used reconstruction-accuracy optimized collapsing (Korhonen et al., 2014) and a resolution of 400 parcels covering the whole cortex.

Cross-talk function and resolution matrix

In MEG/EEG source connectivity studies, a resolution matrix $P = W\Gamma$ ($P \in \mathbb{R}^{n\times n}$) is often used to describe the relationship between true signals and modeled signals from n sources in the absence of noise (Farahibozorg et al., 2017; Hauk and Stenroos, 2014; Hauk et al., 2011; Liu et al., 2002). In P, each diagonal element quantifies the sensitivity for estimating signals from that source. Each row of P is the "cross-talk" function (CTF) that describes the amount of mixing between one source and all other sources. Each column of P is a "point-spread" function (PSFs) that describes how the modeled signal from any one source is spread across all other sources.

The mixing function

For the reconstruction accuracy (fidelity) optimized cortical parcellation (Korhonen et al., 2014), we approximated the resolution matrix P