Identifying causal networks of neuronal sources from EEG/MEG data with the Phase Slope Index - A simulation study

Arne Ewald[a,b,*], Forooz Shahbazi Avarvand[c,d], Guido Nolte[e]

[a] *NIRx Medizintechnik GmbH, Baumbachstr. 17, 13189 Berlin*
[b] *Berlin Institute of Technology, Machine Learning Laboratory, Franklinstr 28/29, 10587 Berlin*
[c] *Fraunhofer FOKUS, Standort Adlershof, Kekulestrasse 7, 12489 Berlin*
[d] *Humboldt-Universitaet zu Berlin, Dept. of Computer Science, Rudower Chausee 25, 10099 Berlin*
[e] *Dept. of Neurophysiology and Pathophysiology, Martinistr. 52, 20246 Hamburg*

Abstract

The investigation of functional neuronal synchronization has recently become a growing field of research. With their high temporal resolution EEG and MEG are well suited measurement techniques to identify networks of interacting sources underlying the recorded data. The analysis of the data in terms of effective connectivity, nevertheless, contains intrinsic issues like the problem of volume conduction and the non uniqueness of the inverse solution. Here we briefly introduce a series of existing methods assessing these problems. To determine the locations of interacting brain sources robust to volume conduction, all computations are solely based on the imaginary part of the cross-spectrum as a trustworthy source of information. Furthermore, we demonstrate the feasibility of estimating causal relationships of systems of neuronal sources with the phase slope index (PSI) in realistically simulated data. Finally, advantages and drawbacks of the applied methodology are highlighted and discussed.

Keywords: Electroencephalography, Magnetoencephalography, Effective Connectivity, Source Localization, Imaginary Part of Coherency, Volume Conduction

*Corresponding author
Email address: mail@aewald.net (Arne Ewald)

1. Introduction

Besides the pure location of neuronal sources, a distinct interaction pattern including different brain regions is hypothesized to determine the function of the brain within a particular task [37, 34, 38, 35]. Whereas the location of brain activity can be well determined with high spatially resolved functional imaging techniques like fMRI, the temporal resolution makes it difficult to capture dynamics inside the brain. In contrast to fMRI, non-invasive measurement techniques like EEG and MEG suffer from a poor spatial resolution. Nevertheless, the temporal resolution in the millisecond range makes them well suitable to study neuronal synchronization which is understood as a mechanism of functional communication (e.g. [8]).

One of the fundamental problems arising for the identification of interacting neuronal sources from EEG or MEG data is the so-called problem of 'volume conduction' or 'field spread' [28]. As the electric (EEG) or magnetic (MEG) field produced by a single source propagates through the whole head, it is captured by at least a couple of sensors on the scalp. Hence, an interaction, determined between two different sensors can arise only due to a single source and does not necessarily reveal information about the underlying network. To overcome this problem, Nolte et. al. proposed to use only the imaginary part of the complex valued coherency as a robust measure of interaction on sensor level [23]. However, the interpretation of relationships between sensors in terms of brain sources is quite difficult in many cases. Therefore, the final aim is to estimate the activity of neuronal sources and then apply appropriate connectivity measures. It is often argued that the procedure of first calculating the source activity also completely solves the problem of volume conduction. But although the effect of volume conduction is most dominant on sensor level, it also affects estimates of source activity, i.e. the determination of synchronized neuronal sources from scalp recordings [33]. One more fundamental issue is that the inverse problem is not uniquely solvable as it is highly underdetermined and,

therefore, mathematically ill-posed. Many different source configurations could give rise to the same EEG/MEG measurement. Therefore, additional information or constraints are required for the determination of the underlying sources [3].

In the present paper, we describe two different methodological procedures to potentially overcome these problems and to identify networks of directed information flow within the brain. They are based on an appropriate pre-processing to diminish the effect of volume conduction and, in addition, to bound the search space for the estimation of interacting neuronal sources. We utilize the properties of the imaginary part of the cross-spectrum (ImCs) by using it as a basis to estimate neuronal source activity. Given the time courses of the interacting brain sources, we finally apply the phase slope index (PSI) to determine the directional coupling between them in order to obtain an interpretable picture of neuronal interaction. In Section 2 an overview about the methodology is given and the individual methods are described briefly. In Section 3, we specify how we simulated EEG data in a realistic fashion and present results of the applied methods to demonstrate the feasibility of the proposed procedure. Finally, results are discussed in Section 4.

2. Methods

In this Section the applied methodology is introduced. In Section 2.1 an overview about all used methods is given followed by a more detailed description of each individual method.

2.1. Overview

A well established method to estimate a linear relationship between two time series in the frequency domain is coherence ('coherence' usually denotes the absolute value of the normalized, complex valued cross-spectrum. To avoid confusion about the terminology we call the complex valued quantity 'coherency'). To assess the problem of volume conduction, it was proposed to focus on the imaginary part of coherency (ImC), as independent

sources do not contribute systematically to the imaginary part of the cross-spectrum [23]. Based on this finding we describe two different ways to estimate networks of brain sources, as shown in Figure 1.

The left branch of this 'methodological tree' starts with the determination of subspaces containing interacting source pairs. To localize pairwise interacting sources from the imaginary part of the cross-spectrum or coherency, respectively, two different methods have been established. The Pairwise Interacting Source Analysis (PISA) decomposes the imaginary part of the cross-spectrum into pairs of interacting sources [25]. It is technically related to standard second order blind source separation methods while the meaning is somewhat opposite as it decomposes only that part of the cross-spectrum which is inconsistent with independent sources. The other method (MaxImC), which is used in the present paper, determines spatial patterns such that the imaginary part of coherency is maximized and, therefore, extracts major large scale interactions. Furthermore, it is independent of the actual mapping from sources to sensors [5]. The result of both methods is a subspace containing the scalp patterns of two interacting sources. These patterns have to be demixed using further assumptions to estimate the truly interacting sources. This is formulated within the Minimum Overlap Component Analysis (MOCA) where linear source estimates are demixed assuming that the true source distribution have minimal spatial overlap. In order to separate interacting sources a spatial rather than a dynamical criterion has to be defined. We make the assumption that separate sources occupy separate brain regions and hence do not overlap spatially. But even if the true sources do not overlap the respective estimated distributed sources in general will. To come as close to the true separation as possible we,therefore, minimize the spatial overlap. This was tested extensively in [19] for dipolar sources reconstructed with weighted minimum norm estimates (WMNE), which was also used here as an inverse method [14].

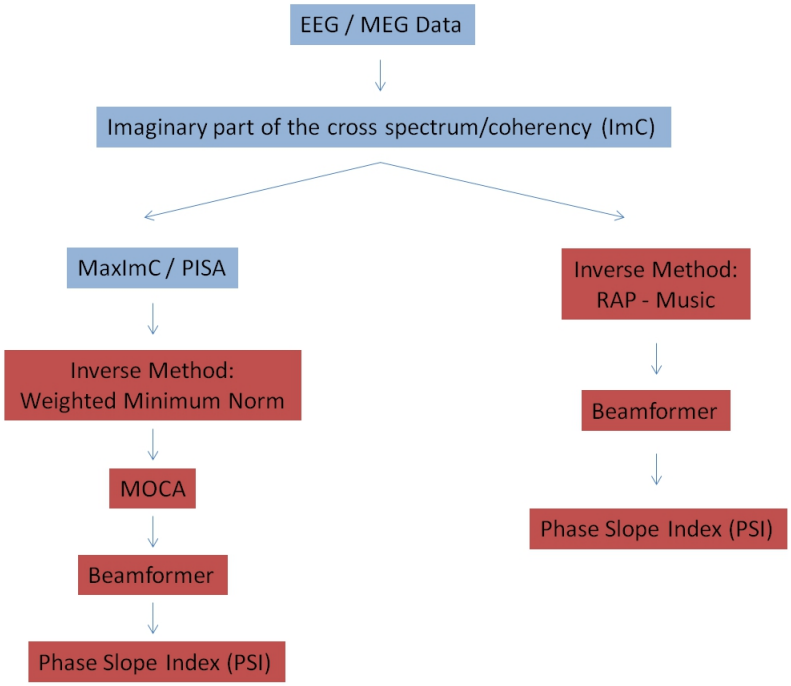

Figure 1: Methodological overview. Two different ways to estimate a causal network of neuronal sources underlying measured EEG or MEG data, robust to artifacts of volume conduction. A blue box indicates a method acting on sensor level and a red box a method in source space.

A different way to estimate interacting sources robust to volume conduction is to apply RAP-MUSIC (see [21]) on a subspace determined by the imaginary part of the cross-spectrum as also shown in [1]. At the end of each methodological branch, a beamformer-like projection is used to determine the time courses of the estimated sources. Finally, the phase slope index (PSI) is applied as an indicator for the direction of the information flow between different time series [27]. In contrast to the original research paper, where PSI is shown on sensor level, we use it here to identify the driver and the recipient from the estimated time series of macroscopic neuronal sources. By analyzing the imaginary part of the cross-spectrum prior to source reconstructions we focus on interacting sources and, hence, diminish artifacts due to non relevant non-interacting or noise sources.

2.2. Robustness of volume conduction of the imaginary part of the cross-spectrum

A key idea of the present paper is to base the localization of interacting sources on the imaginary part of the cross-spectrum (ImCs) as the only available reliable source of

information. Therefore, we want to review the special role of the ImCs shortly. Let us consider the Fourier transform $x_p(f) = r_p e^{i\phi_p}$ in a segment (e.g. an event-related epoch) of measured EEG/MEG data in the sensor p. Then, the complex valued cross-spectrum is defined for each frequency f and for each pair of sensors p and q by

$$C_{pq}(f) = \left\langle x_p(f) x_q^*(f) \right\rangle = \left\langle r_p r_q e^{i(\phi_p - \phi_q)} \right\rangle, \tag{1}$$

where $*$ denotes the complex conjugate and $\langle . \rangle$ describes the expectation value which is usually approximated by averaging over a large number of trials [4]. Furthermore, the signal in an EEG/MEG sensor p can be described as the linear superposition or mixture of K brain sources $s_k(f)$, leading to

$$x_p(f) = \sum_{k=1}^{K} a_{pk} s_k(f). \tag{2}$$

If we assume an instantaneous mapping from sources to sensors, the mixing coefficients a_{pk} are real valued and the signal phases ϕ_p in sensor space are not distorted. This can be derived from the validity of the quasi-static approximation of the Maxwell equations below 2kHz, and therefore in range of EEG/MEG frequencies of interest [31, 36]. Further assuming only independent, i.e. non phase-locked or interacting sources, and plugging Equation (2) into Equation (1), leads to

$$C_{pq}(f) = \sum_{kk'} a_{pk} a_{qk'} \left\langle s_k(f) s_{k'}^*(f) \right\rangle = \sum_{k} a_{pk} a_{qk} \left\langle s_k(f) s_k^*(f) \right\rangle = \sum_{k} a_{pk} a_{qk} \left\langle \left| s_k(f)^2 \right| \right\rangle, \tag{3}$$

which is purely real valued. Hence, independent brain sources are mapped only in the real part of the cross-spectrum and all significant deviation from zero of the ImCs can be interpreted as true brain interaction. Another point of view is that zero-phase interactions are neglected or not interpreted as they are confounded by artifacts of volume conduction. This line of arguments is also valid for complex valued coherency, the normalized cross-

spectrum, as the normalization

$$Coh_{pq}(f) = \frac{\left\langle r_p r_q e^{i(\phi_p - \phi_q)} \right\rangle}{\sqrt{\left\langle r_p^2 \right\rangle \left\langle r_q^2 \right\rangle}} \tag{4}$$

is as well real valued for independent sources.

2.3. Maximizing imaginary coherency

One way of pre-processing to increase signal-noise-ratio in terms of the imaginary part of coherency (ImC) is to determine spatial filters maximizing the ImC [5]. These filters can be converted to spatial patterns (see [22]) that themselves can be interpreted as mixed topographies of the *most dominant* interacting brain sources. Let us consider a pre-whitened imaginary part of the cross spectrum

$$\mathbf{D}(f) = \mathbf{C}_R(f)^{-1/2} \mathbf{C}_I(f) \mathbf{C}_R(f)^{-1/2} \in \mathbb{C}^{N \times N \times F} \tag{5}$$

for all $N \times N$ sensor pairs and F frequencies where $\mathbf{C}_R(f) = \Re(\mathbf{C}(f))$ denotes the real part of the cross-spectrum and $\mathbf{C}_I(f) = \Im(\mathbf{C}(f))$ the imaginary part (for details of the pre-whitening please refer to [5]). Let us furthermore consider the Fourier transform of the data for all N channels $\mathbf{x}(f) = [x_1(f) \dots x_N(f)]^T$ and its whitened form $\mathbf{y}(f) = \mathbf{C}_R(f)^{-1/2}\mathbf{x}(f)$. Then, weights or spatial filters $\mathbf{a} \in \mathbb{R}^{N \times 1}$ and $\mathbf{b} \in \mathbb{R}^{N \times 1}$ can be defined, such that the ImC between the two virtual channels $z_a(f) = \mathbf{a}^T(f)\mathbf{y}(f)$ and $z_b(f) = \mathbf{b}^T(f)\mathbf{y}(f)$ is maximized. The ImC between $z_a(f)$ and $z_b(f)$ can be derived to be

$$ImC_z(f) = \frac{\mathbf{a}^T(f)\mathbf{D}(f)\mathbf{b}(f)}{\|\mathbf{a}(f)\| \, \|\mathbf{b}(f)\|} \tag{6}$$

and maximization of Equation (6) is achieved by solving the eigenvalue equations

$$\mathbf{D}(f)^T\mathbf{D}(f)\mathbf{b}(f) = \lambda^2\mathbf{b}(f) \quad and \quad \mathbf{D}(f)\mathbf{D}(f)^T\mathbf{a}(f) = \lambda^2\mathbf{a}(f). \tag{7}$$

According to the previous derivations, the eigenvectors belonging to the largest eigenvalues of $\mathbf{D}(f)^T\mathbf{D}(f)$ and $\mathbf{D}(f)\mathbf{D}(f)^T$ are the spatial filters $\mathbf{a}$ and $\mathbf{b}$ that maximize the imaginary

part of coherency. These filters could in general be converted into patterns by

$$\tilde{\mathbf{a}}(f) = \mathbf{C}_R(f)\mathbf{a}(f) \quad and \quad \tilde{\mathbf{b}}(f) = \mathbf{C}_R(f)\mathbf{b}(f). \tag{8}$$

as described in [5, 22] and, hence, being interpreted as topographies of interacting sources. But the eigenvalues from Equation (7) are degenerate and occur in complex conjugate pairs. Therefore, the eigenvectors are not unique up to a rotation and every linear combination of the two eigenvectors is an eigenvector itself fulfilling Equation (7). In other words, the calculated topographies in Equation (8) only span a subspace of the real topographies of the underlying brain sources. Therefore additional constraints have to be employed to demix the orthogonal topographies which is addressed in Section 2.4.

A further connectivity measure that can be derived by maximizing the imaginary part of coherency as described above is the Global Interaction Measure (GIM) [5]. The GIM itself is the frequency dependent maximized ImC, i.e. the value obtained in Equation (6), and illustrates at which frequency we observe neuronal synchronization. In this paper, the GIM is used to select the frequency bin of interest from the simulated EEG data.

As an additional remark, we would like to point out another method that determines a subspace of topographies of interacting sources based on completely different assumptions. Pairwise interacting source analysis (PISA) is an adaptation of common blind source seperation techniques like ICA (e.g. [15]) with a focus on interactions [25]. Hence, PISA is also well-suited as starting point to determine a subspace of pairwise interacting sources and for further processing as shown in [26].

2.4. Demixing Sources with MOCA

The two spatial patterns $\tilde{\mathbf{a}}(f)$ and $\tilde{\mathbf{b}}(f)$ from Equation (8) that are obtained by maximizing the ImC, are unique up to a rotation as described in Section 2.3. Therefore, further constraints have to be applied to find a unique representation of underlying source distributions. Here, we use a method called minimum overlap component analysis (MOCA), introduced in [19]. As the idea of minimally overlapping sources can only be implemented

in source space, source distributions belonging to the patterns in Equation (8) have to be estimated. For simplicity, we employ a weighted minimum norm solution (WMN, see e.g. [17] and [16]). The underlying source distribution $\mathbf{s}_{\tilde{a}}$ giving rise to the spatial pattern $\tilde{\mathbf{a}} = \mathbf{A}\mathbf{s}_{\tilde{a}}$ can be estimated by solving

$$\mathbf{s}_{\tilde{a}} = \arg\min_{\mathbf{s}} \|\tilde{\mathbf{a}} - \mathbf{A}\mathbf{s}_{\tilde{a}}\|_2^2 + \lambda \|\mathbf{W}\mathbf{s}_{\tilde{a}}\|_2^2 , \tag{9}$$

with λ being a regularization parameter and $\mathbf{W}$ a weighting matrix, here chosen to penalize deep sources. The matrix $\mathbf{A}$ denotes the lead field that describes the linear mapping from given brain sources to measurement sensors. It is calculated using a realistic volume conductor as described in [24]. Please note, that applying more sophisticated linear inverse solutions might help to improve the performance of using the phase slop index (PSI) on source level. For example, the ℓ_2 norm used in Equation (9) leads to very smooth and, therefore, often too extended source distributions whereas an ℓ_1 norm would generate an often too sparse distribution. To resolve this trade-off, Haufe et. al. have proposed an intermediate measure in [13]. However, an exhaustive discussion on particular inverse solutions is out of the scope of this paper, and, to apply MOCA, it is necessary to use a linear inverse method.

As stated in Section 2.3 the source distributions $\mathbf{s}_{\tilde{a}/\tilde{b}}$ have to be demixed as the respective topographies are unique up to mixing within the respective two-dimensional subspaces. After transforming to spatially uncorrelated source distributions, named $\hat{\mathbf{s}}_{\tilde{a}/\tilde{b}}$, a rotational ambiguitiy remains and the optimally demixed distributions can be expressed as

$$\begin{pmatrix} \mathbf{m}_{\tilde{a}}(j,\varphi) \\ \mathbf{m}_{\tilde{b}}(j,\varphi) \end{pmatrix} = \begin{pmatrix} cos\varphi & sin\varphi \\ -sin\varphi & cos\varphi \end{pmatrix} \begin{pmatrix} \hat{\mathbf{s}}_{\tilde{a}}(j) \\ \hat{\mathbf{s}}_{\tilde{b}}(j) \end{pmatrix} \tag{10}$$

with φ being the rotation angle and j being all brain voxels on a pre-defined grid. Achieving minimum spatial overlap of the source distributions $\mathbf{m}_{\tilde{a}}(j,\varphi)$ and $\mathbf{m}_{\tilde{b}}(j,\varphi)$ can be

realized by analytically minimizing the function

$$O(\varphi) = \sum_j (\mathbf{m}_{\tilde{a}}(j,\varphi)\mathbf{m}_{\tilde{b}}(j,\varphi))^2 \tag{11}$$

defining the overlap [19].

2.5. RAP-MUSIC

A different way of estimating interacting source distributions based on the imaginary part of the cross-spectrum is Recursively Applied and Projected Multiple Signal Classification (RAP Music, [21]), a variant of the MUSIC algorithm [32]. The basic idea behind MUSIC is to define a $p \ll N$ dimensional low-rank subspace projection of data in N measurement channels and an orthogonal noise space. The so called signal subspace $\mathbf{S}$ is usually spanned by the first p eigenvectors of an eigenvalue (or singular value) decomposition of the data's covariance matrix $\overline{\mathbf{C}} \in \mathbb{R}^{N\times N}$. The orthogonal noise subspace is estimated by the span of the remaining $N-p$ eigenvectors of $\overline{\mathbf{C}}$. Given the subspaces, a scan over all pre-defined grid points in the brain is performed to determine whether a source at grid point j is consistent with the signal subspace $\mathbf{S}$ which, for simplicity, is assumed to be defined by normalized and mutually orthogonal columns of $\mathbf{S}$. This consistency can be expressed in terms of the angle ϑ between $\mathbf{S}$ and the forward model, i.e. the projection of a dipolar brain source onto the scalp, at grid point j. If ϑ is small or even zero, a source at grid point j is likely to be contained in the data subspace. With the forward model $\mathbf{L}_j \in \mathbb{R}^{N\times 1}$ the angle ϑ can be defined as

$$\cos^2\vartheta(\mathbf{L}_j,\mathbf{S}) = \frac{\mathbf{L}_j^T\mathbf{S}^T\mathbf{S}\mathbf{L}_j}{\mathbf{L}_j^T\mathbf{L}_j}. \tag{12}$$

Please note that the previous formulation is valid for given dipole orientations. For unknown dipole directions, the forward model $\mathbf{L}_j$ can be expressed by $\mathbf{L}_j = \widetilde{\mathbf{L}}_j\alpha_j$ where $\widetilde{\mathbf{L}}_j$ is an $N\times 3$ matrix for unit dipole directions in x, y and z direction and α_j is 3×1 vector defining the dipole direction at grid point j. Now, the forward model $\mathbf{L}_j$ can be

determined by optimizing over α_j, which can be done analytically [1].

One drawback of the MUSIC algorithm is its failure in the presence of increasing numbers of sources which leads to several maxima for a single scan. As this is the case for interacting sources or even systems of interacting sources, we make use of a variant of the MUSIC algorithm called RAP-MUSIC [21]. Here, the strongest source found in an initial MUSIC scan is projected out and the MUSIC scan is repeated. Then the second strongest source is projected out and so on. This procedure is repeated iteratively for all p sources.

The major modification to RAP-MUSIC that is done for the work presented in this paper is that we do not define the signal subspace in terms of the covariance matrix as stated before. Instead, we apply RAP-MUSIC in the frequency domain on the imaginary part of the cross-spectrum, defined in Equation (1). The reason is to focus on reliable interactions robust to volume conduction and to diminish artifacts from non-interacting sources. Please note, that the cross-spectrum is frequency dependent and hence, calculations in this paper are done for a single frequency. In general, it is also conceivable to average the cross-spectrum over frequencies and, therefore, to apply the proposed methodology in a specific band. However, the determination of sender and recipient of information as described in Section 2.6, is based on a broader frequency range. Therefore, also a distinct frequency band is taken into account for the whole procedure described in this paper.

2.6. The Phase Slope Index

The phase slope index (PSI) is a method to estimate the direction of information flow between two time series [27]. The fundamental concept behind PSI is that in general the cause precedes the effect and interaction is accompanied by a certain time delay τ. Let's consider two time series $\widehat{x}_p(t)$ and $\widehat{x}_q(t)$ where one is the delayed version of the other

$$\widehat{x}_q(t) = c\widehat{x}_p(t-\tau) \tag{13}$$

including an amplification ($c > 1$) or damping ($0 < c < 1$) constant c. With the definition of the cross-spectrum in Equation (1), the relation in Equation (13) of the Fourier transformed signals $x_p(f)$ and $x_q(f)$ leads to

$$C_{pq}(f) = \left\langle x_p(f) x_q^*(f) \right\rangle = \left\langle r_p^2 c \, e^{i\,2\pi f \tau} \right\rangle \sim e^{i\,2\pi f \tau} \equiv e^{i\phi(f)}. \tag{14}$$

From Equation (14) we can observe that the phase spectrum

$$\phi(f) = 2\pi f \tau \tag{15}$$

itself is linearly dependent on frequency and proportional to the time delay τ. Therefore, a positive slope of the function $\phi(f)$ indicates a positive τ and according to the example in Equation (13) an information flow from $\widehat{x}_p(t)$ to $\widehat{x}_q(t)$. A negative slope and, hence, a negative τ would indicate a directed information flow from $\widehat{x}_q(t)$ to $\widehat{x}_p(t)$. Including further requirements, such as statistical robustness and insensitivity to non-interacting signal parts (see [27]), the final formulation of PSI is given by

$$\Psi(f) = \Im(\sum_{f \in F} Coh_{pq}^*(f) Coh_{pq}(f + \delta f)) \tag{16}$$

where $Coh_{pq}(f)$ is the complex coherency as defined in Equation (4) and δf is the frequency resolution in the frequency band F in which the phase slope is estimated. As a reasonable property, the value for PSI in Equation (16) fluctuates around zero. Including the estimation of the standard deviation, e.g. with a Jackknife procedure [20], the significance level of PSI can be evaluated. Hence, a result within a certain confidence interval around zero would be neglected as not interpretable. In this way, robustness and reliability of the method is enhanced.

As evaluating the direction of information flow with PSI is based on temporal assumptions, the time series of the individual sources have to be determined. This can be achieved by projecting the measured sensor data onto the topographies of the calculated

sources. This procedure is formally equivalent to a beamforming with spatially white noise.

3. Results

In this Section the results of the simulations are presented. In Section 3.1 it is described how data were simulated to reveal EEG/MEG properties. Furthermore, the result of first maximizing the ImC, then applying MOCA to demix topographies and finally, using PSI to estimate the direction of information flow is illustrated. In Section 3.2 the simulation of an interacting system of four sources is demonstrated and how RAP-MUSIC and PSI are applied to determine the locations of neuronal sources and the causal relationships among them.

3.1. Maximizing the imaginary coherency, MOCA and PSI

To realistically simulate EEG data of two interacting sources, we generated random data according to an AR-model of order 10, with 60000 time points and additional noise of 20 percent of the signal power. All Coefficients of the AR-model were randomly chosen but coefficients on the respective off diagonal were set to zero such that the second time course was simulated to be driven by the first one. The cross-spectrum was obtained by segmenting the data in 512 data points long epochs and performing an FFT on the Hanning windowed data. As the calculations described in Section 2 are performed in a specific band in the frequency domain, we assured that minimum 30 percent of the signal power of the two simulated time courses is contained in a specific band. This band, i.e. the most dominant frequency bin, is then automatically selected with the GIM for further processing (see Section 2.3). In general, the suggested procedure can be applied in any frequency range below 2kHz and, therefore, in any band relevant for EEG/MEG analysis. Although it is common in practice to analyze oscillations in the alpha range (9-13Hz) or the beta range (17-25Hz) as strong brain oscillation occur in these bands, the investigation of any EEG/MEG relevant band is conceivable as long as a prominent oscillatory signal

is detectable. Hence, the choice of frequency in the presented simulations is arbitrary and only the presence of a signal in a specific band matters.

In addition to the dynamics, source locations were defined by two dipoles, one in each hemisphere. As a head model we used a standardized MNI head obtained from an average of 152 subjects [6, 7]. According to the previous definition of the time courses, the source in the left hemisphere drives the source in the right hemisphere. These resulting time courses were projected to sensor space (59 EEG channels) by the randomly mixed patterns of the two source dipoles. Again, noise was added with 10% of the size of the simulated EEG signal. Figure 2 shows the simulated dipoles, the resulting EEG topographies for 56 sensors, the mixed topographies, the imaginary part of coherency for each channel pair, and the GIM of the modeled source data.

One of the main motivations to apply measures of effective connectivity on *source* level is the interpretability of the results. Figure 3 shows the results of the phase slope index between each pair of sensors. At a particular frequency, here the one selected with the GIM (see vertical line in Figure 2.D), these bivariate connectivities can be visualized in a so called head in head plot. Each small circle inside the big schematic scalp shows the connectivity of this particular EEG electrode to all other electrodes. For orientation purposes, a small black dot is shown inside each small circle again indicating the position of the particular reference electrode on the scalp. A cold color and a negative value of PSI shows that the particular measurement channel receives information from a distinct recording site whereas a warm color indicates that the channel is sending information. One can observe that the result does not clearly reflect the underlying simulated source structure. Even with simple interaction schemes these head-in-head plots are not easy to interpret in terms of interacting brain sources. Furthermore, results on sensor level always depend on the choice of reference which may distort locations of brain regions on sensor level [12].

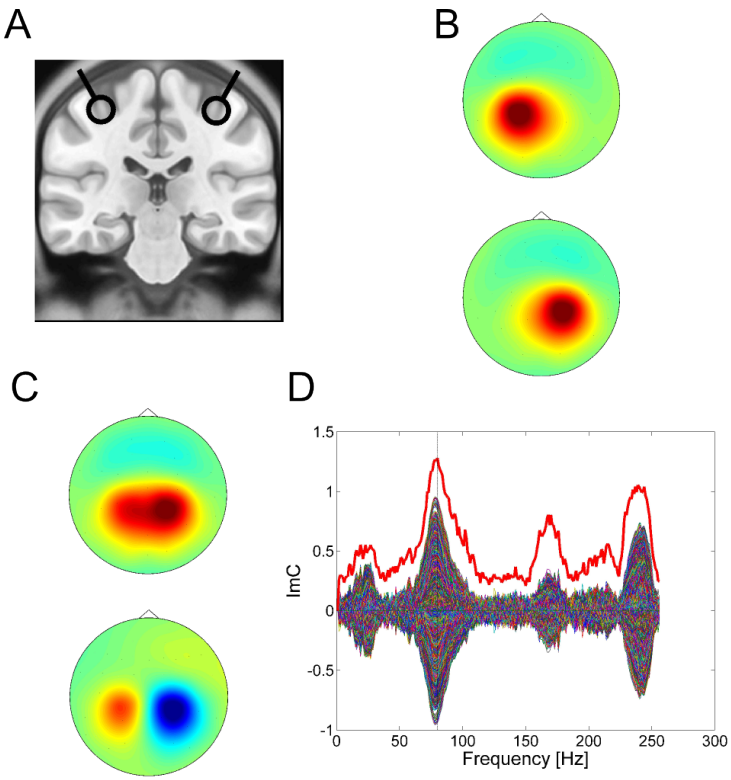

Figure 2: The simulated system of two interacting sources. A: Two dipoles that are simulated in different hemispheres. For displaying purposes only the relevant MRI slice is shown. Data was simulated such that the source on the left drives the sources on the right. B: Topographies of the underlying sources. C: Artificial mixture of the two topographies with a random mixing matrix. As the data were scaled by the mean of the signal power, the scale of the topographies is irrelevant for this simulation and color bars are neglected. D: The imaginary part of coherency for each channel pair and GIM over frequency. By choosing the maximum value of the GIM a particular frequency band (or a single bin) of interest is selected.

The results of the source localization and demixing can be found in Figure 4. Based on the spatial patterns that are obtained by maximizing the imaginary part of coherency (see Section 2.3), the underlying source distributions are calculated with a minimum norm estimator and shown in Figure 4.A. Comparing with the initially simulated dipoles, one can observe that the sources are not separated properly. This problem is addressed with MOCA. The results in Figure 4.B demonstrate that the sources are being well demixed in the present example and match the originally defined dipoles.

Given the sources, the time courses were estimated by projecting onto the source topographies and PSI was calculated for the two time series. To estimate PSI, we used

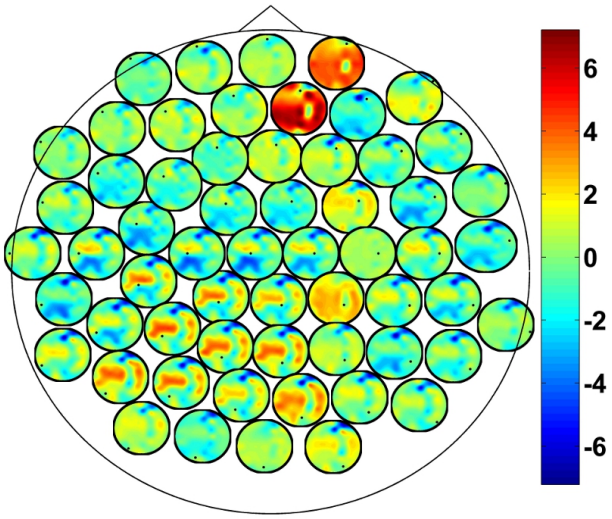

Figure 3: The results of the phase slope index on sensor level

the whole frequency spectrum of the data. Furthermore, the standard deviation was approximated with a Jackknife estimator. We assume that $PSI/StdDev$ is approximately Gaussian distributed with unit standard deviation. Then a p value of .05 corresponds to $|PSI/StdDev| > 1.96$ which was approximated by 2 for simplicity. Even though this is not exactly true we consider this as reasonable. In Figure 7 we show results of a simulation with 4% false detections for mixtures of sources, which is formally equivalent to zero delay, indicating that our approach is slightly over-conservative. Table 1 shows the results for this particular simulation case. A positive PSI of 0.4 indicates an information flow from the first source (left one in Figure 4.B) to the second source (right one in Figure 4.B) which resembles the way the data was simulated. The standard deviation is about an order of magnitude smaller that the value for PSI itself. Hence, we would consider the result as significant which is also indicated by the ratio of PSI and its standard deviation.

Table 1: The phase slope index and its standard deviation for the example of two interacting sources.

Phase Slope Index (PSI)	Standard Deviation	PSI/StdDev
0.39695	0.047187	8.4124

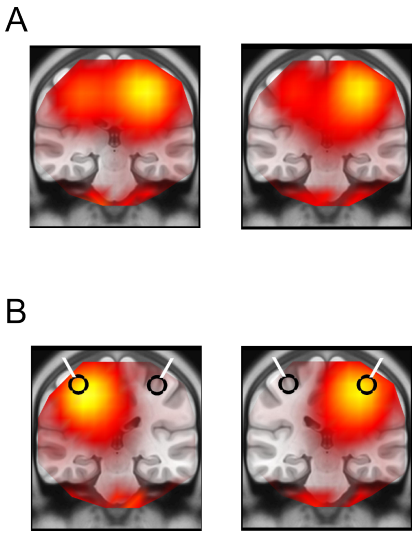

Figure 4: Source localization and demixing. A: The resulting sources obtained with a weighted minimum norm solution on the basis of the topographies found by maximizing the imaginary part of coherency. B: The sources demixed with MOCA and the dipoles that have been initially simulated.

In an additional simulation we have varied the generation of noise. Here, only a single time course is modeled by an AR model to generate data with a distinct frequency component. The time course of the second source is obtained by shifting the first source by 4 data points. Now, noise was randomly generated for each voxel inside the brain and projected onto the scalp. Thus, way more noise sources are present than brain sources. Finally, data and noise were normalized with their mean power and added. To investigate the behavior of the proposed processing scheme, we run the simulation N=1300 times with randomly chosen source dipole locations and orientations inside the brain. The results are shown in Figure 5.

The upper left plot in Figure 5 shows the result of PSI divided by its standard deviation over the source localization error for each run. If the run showed the correct causal information flow is color-coded. Depending on which source is found as the first source by the source localization procedure, PSI can either be negative or positive which can be observed from the two centroids of the point clouds. To judge the correctness of a run and to assign the sources, the distances between the modeled and the estimated sources

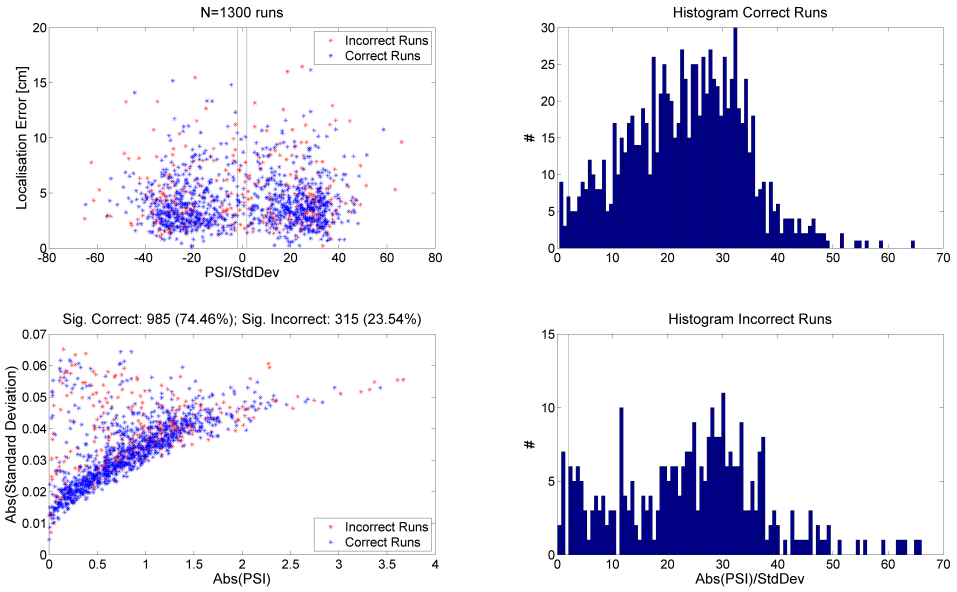

Figure 5: The results of the statistics for PSI between the two modelled sources with a constant phase shift of 4 data points.

have been calculated and minimized.

From the N=1300 runs, 74.46% turned out to be significantly ($|PSI/StdDev| > 2$) correct and 23.45% significantly incorrect. The fairly large number of incorrect results is a consequence of mislocalizations. Apparently, even if an estimated source is closer to, say, the first true source, it is in general possible that it picks up more activity from the second true source. Specifically, the randomly assigned sources could have been located too close to each other such that MOCA's assumptions of minimal spatial overlap does not hold for respective source estimates. If the source locations remain fixed and well separated, i.e. as in the previous simulation shown in Figure 4, the described methodology always returns the correct result. Figure 6 shows the histogram for this simulation.

As a kind of sanity check for the statistical properties of the imaginary part of the cross-spectrum, we have executed the same simulation but with zero phase delay between the modelled sources. As expected, PSI returned mostly no significant result as shown in Figure 7.

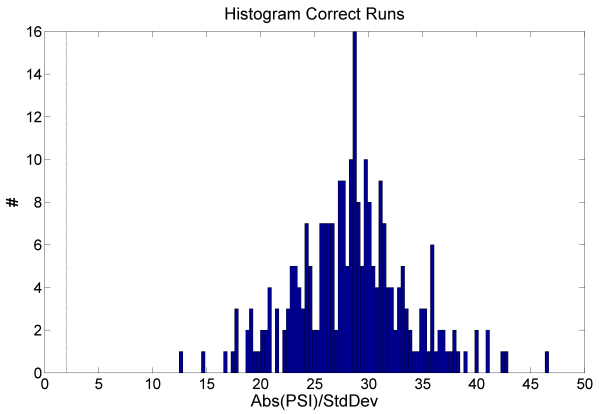

Figure 6: Histogram for PSI over the standard deviation with two causally connected sources modelled in somatosensory areas as shown in Figure 8.A

3.2. RAP-MUSIC and PSI

A further simulation consists of a system of four interacting sources. The location of the sources, again modeled as dipoles, are shown in Figure 8.A. Time courses of the individual brain sources are simulated by an autoregressive model of order 10 with the same noise structure as described in Section 3.1. The information flow goes from the left source to the second left source, from the second left source to the second right source and from the second right source to the right source. Hence, the source on the right only receives information.

The EEG data was simulated the same way as before: The modeled time series were projected to 59 EEG sensors by the topographies of the four sources. Again, noise was added. Figure 8.B shows the imaginary part of coherency for all pairs of sensors. Furthermore, the GIM is shown and used to determine the frequency of interest, i.e. the frequency bin where the most prominent interaction is present. A couple of peaks are visible that exceed the present noise level and the one with the maximum GIM is selected for further processing.

Figure 8.C illustrates the phase slope index (divided by its standard deviation) on sensor level at the chosen frequency bin. The plot shows that there is significant directed interaction present in the data. Concerning the location of the underlying interaction one

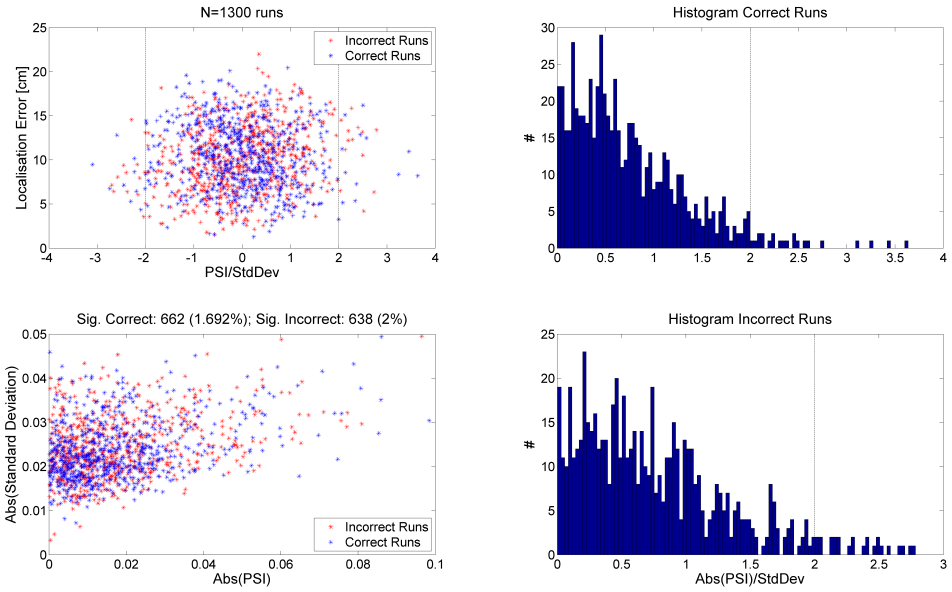

Figure 7: Statistical results for two modeled brain sources with zero phase delay. Almost no statistical significant result is obtained due to the properties of the imaginary part of the cross-spectrum.

would interpret this head-in-head plot as information passing from frontal to occipital brain areas. According to the simulated sources and the simulated information, the picture provided on sensor level is not correct. As in the example before, the need for reliable calculations in source space is motivated.

Figure 9 shows the results of the RAP-MUSIC scan. Please note, that results from a MUSIC scan do not necessarily represent source *distributions* as for every voxel

$$\frac{1}{1-\cos^2(\vartheta)} \tag{17}$$

(ϑ beeing the angle between a source at a particular voxel and the data subspace, see Section 2.5) is plotted and color coded. However, by plotting the results for all voxels and not only the maximum, one can judge the quality of the source reconstruction. If, for instance, a source would be distributed through the whole head and no clear maximum is visible, the results would be questionable. In the example shown in Figure 9 one can observe that especially for sources '1' and '4' the found locations coincide almost perfectly

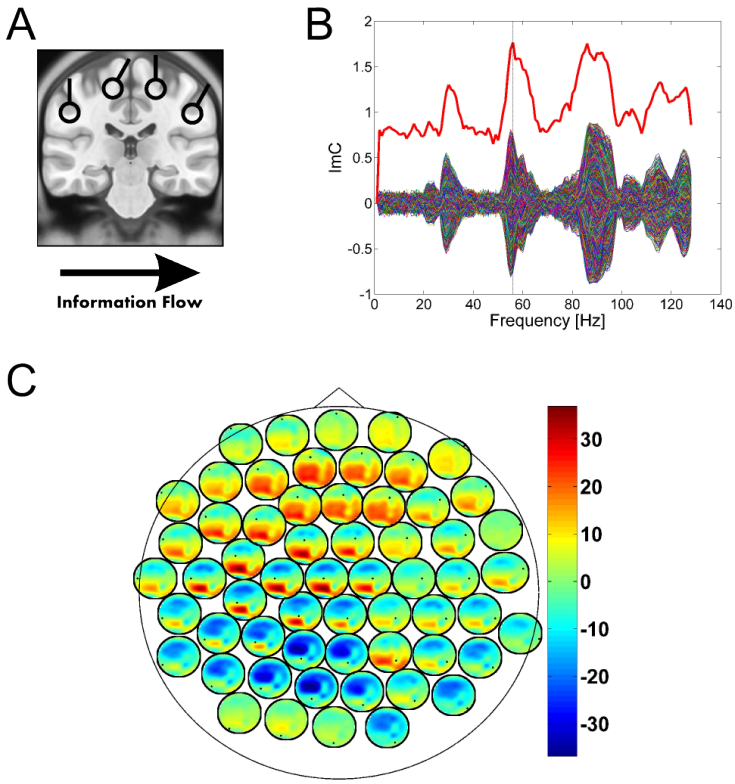

Figure 8: Simulated system of four interacting sources. A: The location of the sources simulated as dipoles. The information flow is modeled from left to right, i.e. the left source sends to the next one on the right and so on. The source on the right only receives information from its neighboring left one. B: The imaginary part of coherency for all pairs of sensors and the selection of frequency with the global interaction measure (GIM, red line). C: The bivariate phase slope index on sensor level visualized as a head in head plot at the frequency selected with GIM.

with the previously modeled dipoles. For sources '2' and '3' it seems that the sources could not be completely demixed, i.e. projected out in the process of the RAP-MUSIC iteration (see Section 2.5). However, maxima are found close to the locations of the modeled dipoles. We used the resulting dipoles of the RAP-MUSIC scan (as stated in Section 2.5, dipole orientations are found by an optimization) to estimate the time series at these four locations with an appropiate projection of the measured sensor data to source space.

Having calculated the time series of the individual sources found in the RAP-MUSIC scan, the phase slope index was evaluated as a bivariate measure between all sources. Table 2 shows the value of PSI over its standard deviation estimated with a Jackknife procedure for all combinations of sources. The sources in the rows serve as references and

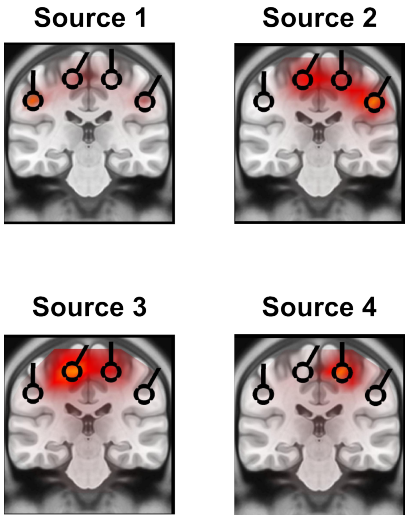

Figure 9: The four sources found with RAP-MUSIC as color coded "distributions". To compare, the originally simulated dipoles are also displayed.

denote the coupling of the particular source to all other sources listed in the columns. For example, a positive value between source '1' (row) and source '2' (column) is interpreted as source '1' being the driver and source '2' being the recipient of information flow between these two sources. Please note, that the result table shows an antisymmetric structure: If coupling between source '1' and '2' two is positive, the coupling between source '2' and '1' has to be equal in magnitude and negative.

The results shown in Table 2 reflect the dynamics that have been simulated. Source '1', the most left one (see Figures 8 and 9) is sending to its neighboring one on the right. As information is passed further to the right, source '1' is sending to all other sources. This effect is also visible as the first row in Table 2 only has positive values. According to that, source '2', the one on the very right, receives information from all other sources, and the second row only has negative values. The finding of the other sources are also in line with the simulated dynamics: Source '3' receives information (negative value) from source '1' and sends (positive value) to source '2' and '4'. Finally, source '4' receives from source '1' and source '3' and sends to source '2'. One can also observe from Table 2 that all values are highly significant concerning the ratio of PSI and its standard deviation. In

practice, it is suggested to consider a value of $|PSI/StdDev| > 2$ as significant [27].

Table 2: The phase slope index over its standard deviation for the simulated system of four interacting sources.

PSI/StdDev	Source 1	Source 2	Source 3	Source 4
Source 1	0	30.80	28.02	36.35
Source 2	-30.80	0	-54.17	-68.53
Source 3	-28.02	54.17	0	46.10
Source 4	-36.35	68.53	-46.10	0

4. Discussion

In the present study we introduced a combination of existing methods to estimate directed coupling between neuronal sources from EEG or MEG data. The focus of the applied processing scheme lies on reliability which is addressed in three different ways. First, artifacts of volume conduction are avoided by using the imaginary part of the cross-spectrum as a reliable basis for all further calculations. Second, subspace methods are used to infer additional constraints for bounding the non-unique inverse problem. Third, a method for calculating the directionality of information flow between brain sources is used that incorporates the estimation of the standard deviation and the definition of a confidence interval. Hence, certain findings of directionality can be easily neglected as not interpretable [27]. Please note, that the central concept of causality behind PSI is that the cause temporally precedes the effect. Hence, a signal feature observed earlier in signal 'A' and later in signal 'B' would lead to classify 'A' as the driver. This temporal argument doesn't prove causality and one can construct counterexamples. However, using random dynamical systems it can be seen that such counterexamples are extremely rare, and we therefore consider a significant PSI as a strong argument for a causal relation.

In particular, we showed how to determine causal relationships between two sources in a distinct frequency band that are obtained by maximizing the imaginary part of

coherency. This procedure is generally extendable to more than two sources by using more pairs of eigenvectors obtained by the maximization. However, this methodology is bounded to determine pairs of interacting sources and no entire systems. As an inverse method for this methodological approach we used a weighted minimum norm estimator and the results shown are rather too smooth distributions centered around the modeled dipole. Here we see some room for improvement by applying more sophisticated linear inverse solutions. However, MOCA was able to demix the overlapping sources in the given example. The question remains if MOCA introduces a bias towards remote interactions. For very localized interactions the assumption of spatially non overlapping sources may be violated. In combination with different inverse methods, e.g. based on the ℓ_1 norm, this problem needs to be evaluated which is out of the scope of this paper. After estimating time courses of the sources, the directionality of coupling was determined correctly and significantly by PSI. Additionally, we have performed statistics to judge the performace of the proposed methodology with a different noise structure. In contrast to dominatly correct results on source level, the picture provided for bivariate connections on sensor level was rather fuzzy and not interpretable in terms of brain sources. The comparison between source and sensor level illustrates the urgent need for the application of analysis methods on source level to obtain a clear picture if interacting brain sources.

The second approach also supports this finding. Here, RAP-MUSIC applied on a subspace based on the imaginary part of the cross-spectrum was used to determine the sources. Although RAP-MUSIC was not able to separate all four sources perfectly, the estimation of causal relationships worked out accuratly. It seems that RAP-MUSIC can be improved especially in the context of interacting sources which is an ongoing research subject. A further issue is the definition of the numbers of sources for RAP-MUSIC that has to be defined in advance. To our knowledge no feasible algorithm has been discovered yet to answer this question. However, the causal relationships between more than two

interacting sources can be estimated reliably by applying RAP-MUSIC on the ImCs and PSI.

The aim of the study is to investigate the feasibility of applying the phase slope index in source space and to use the imaginary part of the cross-spectrum as a trustworthy and fundamental information source for EEG/MEG connectivity analysis. To further evaluate the practical applicability of the whole procedure it would be necessary to evaluate the performance of both presented approaches in order to compare them with other existing techniques such as (s/e)LORETA, DTF, PDC, DICS, Granger causality, DCM, combinations among them and with combinations of methods used in this paper [29, 30, 18, 2, 11, 10, 9]. Furthermore, the behavior of the presented approach in the presence of more noise or different noise structures needs to be investigated and evaluated by further statistics. In a final step, the proposed methodology needs to be evaluated on real data with a known underlying causal structure of known brain sources. This can only be achieved in comparison with invasively recorded data at relevant brain sites or even in the entire brain.

References

[1] F. S. Avarvand, A. Ewald, and G. Nolte. Localizing true brain interactions from EEG and MEG data with subspace methods and modified beamformers. *Comp. Math. Methods in Medicine*, 2012, 2012.

[2] L. Baccala and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.*, 84:463–474, 2001.

[3] S. Baillet, J. Mosher, and R. Leahy. Electromagnetic brain mapping. *Signal Processing Magazine, IEEE*, 18(6):14 –30, Nov. 2001.

[4] J. Bendat and A. Piersol. *Random data; analysis and measurement procedures*. Wiley-Interscience New York,, 1971.

[5] 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*, 60:476 – 488, 2012.

[6] V. Fonov, A. Evans, R. McKinstry, C. Almli, and D. Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, 2009.

[7] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, and D. L. Collins. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313 – 327, 2011.

[8] P. Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *TRENDS in Cognitive Sciences*, 9(10):474–480, October 2005.

[9] K. J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, Aug 2003.

[10] C. Granger. Investigating causal relations by economic models and cross-sprectal methods. *Econometrica*, 37:424–438, 1969.

[11] J. Gross, J. Kujala, M. Hamalainen, 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*, 98(2):694–699, Jan. 2001.

[12] S. Haufe, V. Nikulin, K. Müller, and G. Nolte. A critical assessment of connectivity measures for EEG data: A simulation study. *NeuroImage*, 2012.

[13] S. Haufe, V. V. Nikulin, A. Ziehe, K.-R. Müller, and G. Nolte. Combining sparsity and rotational invariance in EEG/MEG source reconstruction. *NeuroImage*, 42(2):726 – 738, 2008.

[14] M. Hämäläinen and R. Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical and Biological Engineering and Computing*, 32:35–42, 1994. 10.1007/BF02512476.

[15] A. Hyvarinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Netw*, 13(4-5):411–430, 2000.

[16] A. A. Ioannides, J. P. R. Bolton, and C. J. S. Clarke. Continuous probabilistic solutions to the biomagnetic inverse problem. *Inverse Problems*, 6(4):523, 1990.

[17] L. R. Jeffs, B. and M. Singh. An evaluation of methods for neuromagnetic image reconstruction. *IEEE Trans Biomed Eng*, 24(0018-9294 (Linking)):713–723, 1987.

[18] 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. *Biol Cybern*, 85(2):145–157, Aug 2001.

[19] L. Marzetti, C. D. Gratta, and G. N. G. Understanding brain connectivity from EEG data by identifying systems composed of interacting sources. *Neuroimage*, 42(1):87–98, 2008.

[20] R. G. Miller. The jackknife-a review. *Biometrika*, 61(1):1–15, 1974.

[21] J. Mosher and R. Leahy. Source localization using recursively applied and projected (rap) music. *Signal Processing, IEEE Transactions on*, 47(2):332–340, 1999.

[22] V. V. Nikulin, G. Nolte, and G. Curio. A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition. *NeuroImage*, 55(4):1528 – 1535, 2011.

[23] 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*, 115(10):2292 – 2307, 2004.

[24] G. Nolte and G. Dassios. Analytic expansion of the eeg lead field for realistic volume conductors. *Physics in Medicine and Biology*, 50(16):3807, 2005.

[25] G. Nolte, F. Meinecke, A. Ziehe, and K. Müller. Identifying interactions in mixed and noisy complex systems. *Phys Rev E*, 73:051913, 2006.

[26] G. Nolte and K. Müller. Localizing and estimating causal relations of interacting brain rhythms. *Frontiers in Human Neuroscience*, 4(00209), 2010.

[27] G. Nolte, A. Ziehe, V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and K. Müller. Robustly Estimating the Flow Direction of Information in Complex Physical Systems. *Physical Review Letters*, 100:234101, June 2008.

[28] P. Nunez, R. Srinivasan, A. Westdorf, R. Wijesinghe, D. Tucker, R. Silberstein, and P. Cadusch. EEG coherency. i. statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.*, 103:499–515, 1997.

[29] R. D. Pascual-Marqui. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol*, 24 Suppl D:5–12, 2002.

[30] R. D. Pascual-Marqui, D. Lehmann, M. Koukkou, K. Kochi, P. Anderer, B. Saletu, H. Tanaka, K. Hirata, E. R. John, L. Prichep, R. Biscay-Lirio, and T. Kinoshita. Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philos Transact A Math Phys Eng Sci*, 369(1952):3768–3784, Oct 2011.

[31] H. D. Plonsey, R. Considerations of quasi-stationarity in electrophysiological systems. *Bulletin of Mathematical Biophysics*, 29:657–664, 1967.

[32] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 32:276 – 280, 1986.

[33] J. Schoffelen and J.Gross. Source connectivity analysis with MEG and EEG. *Hum Brain Mapp.*, 30(6):1857–65, 2009.

[34] W. Singer. Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24(1):49–65, Sept. 1999.

[35] W. Singer and C. M. Gray. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.*, 18:555–586, 1995.

[36] J. Stinstra and M. Peters. The volume conductor may act as a temporal filter on the ecg and EEG. *Med Biol Eng Comput*, 36(6):711–6, 1998.

[37] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie. The brainweb: Phase synchronization and large-scale integration. *Nat Rev Neurosci*, 2(4):229–239, Apr. 2001.

[38] S. Zeki and S. Shipp. The functional logic of cortical connections. *Nature*, 335(6188):311–317, Sep 1988.