Single-trial connectivity estimation for classification of motor imagery data

Martin Billinger, Clemens Brunner and Gernot R Müller-Putz

Institute for Knowledge Discovery, Graz University of Technology, Graz, Austria
E-mail: clemens.brunner@tugraz.at

Received 26 March 2013
Accepted for publication 24 May 2013
Published 11 June 2013
Online at stacks.iop.org/JNE/10/046006

Abstract

Objective. Many brain–computer interfaces (BCIs) use band power (BP) changes in the electroencephalogram to distinguish between different motor imagery (MI) patterns. Most current approaches do not take connectivity of separated brain areas into account. Our objective is to introduce single-trial connectivity features and apply these features to BCI data.

Approach. We introduce a procedure for extracting single-trial connectivity estimates from vector autoregressive (VAR) models of independent components in a BCI setting.

Main results. In a simulated BCI, we demonstrate that the directed transfer function (DTF) with full-frequency normalization and the direct DTF give classification results similar to BP, while other measures such as the partial directed coherence perform significantly worse.

Significance. We show that single-trial MI classification is possible with connectivity measures extracted from VAR models, and that a BCI could potentially utilize such measures.

1. Introduction

A brain–computer interface (BCI) is a communication device that classifies brain activity and controls a device such as a spelling application, a neuroprosthesis, or a wheelchair [1]. Most non-invasive BCIs rely on the electroencephalogram (EEG) to record brain signals. A typical signal processing pipeline for such a BCI comprises preprocessing, feature extraction, and classification stages. In the preprocessing stage, signals are typically filtered in the spatial and/or spectral domain. Spatial filters usually create a linear mixture of existing signal channels; popular techniques include bipolar derivations, Laplace filters, and common spatial patterns (CSP) [2]. Next, the preprocessed data is further processed in the feature extraction stage. To date, most non-invasive BCIs have used features derived from individual channels. Commonly used feature types include band power (BP), autoregressive (AR) model coefficients, and wavelets [3]. Using suitable classification algorithms in the next stage, these features allow BCIs to discriminate between different brain states or mental tasks. However, these features do not contain information about causal relationships between channels. It is reasonable to assume that such information could provide useful features for BCIs.

The interaction of spatially separated brain areas can be observed through functional or effective connectivity [4, 5]. Functional connectivity reflects statistically related activation of brain areas, while effective connectivity explains the causality of these observed dependencies [5]. Numerous connectivity measures have been developed and applied in neuroscientific studies, including coherence (COH), partial directed coherence (PDC) [6], directed transfer function (DTF) [7], and many others [8, 9]. Some of these measures have also been used in BCI studies [10–18]. Most of these measures are related to vector autoregressive (VAR) models and can be extracted from the model coefficients.

Volume conduction or instantaneous connectivity causes spurious connectivity between raw EEG channels [19]. Different solutions to this problem have been proposed. These solutions can be divided into three groups: (1) connectivity measures designed to suppress instantaneous effects [20–22], (2) inclusion of an instantaneous term in VAR models [23–25], and (3) modeling the EEG as a mixture of sources [26, 27].
We decided to tackle the problem of instantaneous connectivity by applying a VAR model to EEG sources. In contrast to measures specifically designed to suppress volume conduction, our procedures can be easily extended to include new connectivity measures, provided that these new measures can be extracted from VAR models.

Adding an additional instantaneous term in the VAR model definition, as suggested in [23], imposes certain restrictions on the structure of this term. Specifically, the instantaneous term cannot contain any cycles [25]. This implies that instantaneous interaction in any pair of signals can originate from only one of these signals. In other words, if the signal spreads from electrode A to B, it cannot spread from B to A.

In contrast, modeling the EEG as a mixture of sources does not impose such restrictions. Instead, sources are assumed to be independent and non-Gaussian, which is a reasonable assumption for EEG sources [28]. This assumption makes blind source identification possible with independent component analysis (ICA). Since these sources are independent, they exhibit minimal instantaneous interaction. Therefore, they can be modeled with simple VAR models.

For our purposes, modeling the EEG as a mixture of sources results in a better model than adding an instantaneous term, because EEG is commonly assumed to be the result of mostly cortical sources that appear smeared over the scalp due to volume conduction [28]. Thus, we use the connectivity estimation approach recommended and implemented in EEGLAB with the SIFT toolbox [27]: fitting of VAR models to component signals obtained by extended Infomax ICA [29].

For use in BCIs, connectivity measures need to be extracted online from the on-going EEG. However, in practice, many connectivity measures are extracted from VAR models, which are fitted to data that contains multiple realizations of the EEG time series (multiple trials) [30]. Single-trial estimation is inherently more difficult, because only a fraction of the data is available for model fitting.

In this paper, we propose a procedure for obtaining single-trial estimates of connectivity measures between ICA sources for motor imagery (MI) classification. We solve the problem of single-trial estimation by automatically selecting a subset of signals for analysis to reduce the amount of data. This procedure enables us to evaluate different measures in a BCI simulation. The measures we test include univariate measures as well as functional and effective connectivity measures for classification of BCI tasks. Although we use pre-recorded data in our analysis, our work flow easily generalizes to online applications.

2. Methods

2.1. Data acquisition

We recorded 45 EEG and three electrooculogram (EOG) channels with sintered Ag/AgCl ring electrodes. The locations of the EEG channels corresponded to the extended international 10–20 system (see figure 1 for montage details). We recorded all signals at a sampling rate of 300 Hz with three synchronized g.USBamp amplifiers (g.tec, Guger Technologies OEG, Graz, Austria). The amplifiers filtered the raw data with a 0.5–100 Hz bandpass and a 50 Hz notch filter. EOG components in the EEG were reduced with a regression-based approach [31]. For this study, the signals were further resampled to 100 Hz.

Fourteen healthy volunteers (five male, age 29.6 ± 4.2 and nine female, age 24.8 ± 2.3) without prior experience in BCI control participated in our BCI experiment. All participants except one female were right-handed. Each participant took part in two sessions on separate days, with six screening and three feedback runs on each day. At the beginning of each session, participants performed an artifact run to estimate the influence of EOG artifacts on the EEG. The screening runs were based on the synchronous Graz BCI training paradigm [32], which we modified to visually resemble the feedback runs in this study. We only use data from the artifact and screening runs in this study. During feedback runs, we instructed the participants to steer a virtual plane along a path using hand and foot MI. After four screening runs, one feedback run, two further screening runs, and two final feedback runs followed.

Each session comprised 90 right hand MI and 90 foot MI trials. Figure 2 illustrates the timing of a trial. A trial lasted 7 s, followed by a break of 3 ± 0.5 s. An acoustic beep and the appearance of a fixation symbol indicated the start of a trial. 2.5 s after the start of the trial, an arrow appeared, pointing either up or down to instruct participants to perform hand or foot MI, respectively.
2.2. Autoregressive models for connectivity estimation.

We can derive various connectivity measures from VAR models, which are a multivariate generalization of univariate AR models. VAR models consist of univariate and multivariate terms. Univariate terms model individual channels, whereas multivariate terms describe how channels depend on past values of other channels. As such, VAR models are related to the concept of Granger causality. Their structure mirrors effective connectivity in the time domain, as shown in (1), where the \( [M \times 1] \) vectors \( y[n] \) and \( x[n] \) are the \( M \)-channel EEG sample and residual at time instant \( n \), and \( \beta^{(k)} \) is the \( [M \times M] \) AR coefficient matrix at time lag \( k \). By transforming the model equations into the frequency domain, we obtain (2), a spectral representation of the model. The \( [M \times 1] \) vectors \( y(z) \) and \( x(z) \) correspond to EEG and residuals in the \( z \)-domain.

We define the \( [M \times M] \) matrix function \( A(z) \) in (3), which is the inverse of the multivariate transfer function \( H(z) \), as shown in (4):

\[
y[n] = \sum_{k=1}^{p} \beta^{(k)} y[n-k] + x[n] \tag{1}
\]

\[
x(z) = \left( I - \sum_{k=1}^{p} \beta^{(k)} z^{-k} \right) y(z) \tag{2}
\]

\[
A(z) = I - \sum_{k=1}^{p} \beta^{(k)} z^{-k} \tag{3}
\]

\[
y(z) = A(z)^{-1} x(z) = H(z) x(z). \tag{4}
\]

2.3. Instantaneous connectivity

Applying VAR models directly to raw EEG data is problematic. Volume conduction causes EEG channels to be highly correlated. This is expressed as instantaneous connectivity between EEG channels, which is not included in the VAR model. Instead, the residuals \( x \) are no longer independent; they have non-zero covariances. Some connectivity measures such as the PDC or DTF do not include the residual covariance matrix in their definition. Thus, they are distorted by instantaneous connectivity.

We propose a solution based on blind source separation, which does not impose directional restrictions as is the case when adding an instantaneous term. It is possible to use ICA as a preprocessing step to transform the EEG from the electrode space into the source space. ICA minimizes the statistical dependence between signals. We use the extended Infomax ICA algorithm [29] to minimize instantaneous dependencies. Delayed dependencies between signals are preserved and can thus be modeled with VAR models.

In this paper, we refer to data channels obtained from ICA as sources, although they are not necessarily related to cortical sources or dipoles.

2.4. Connectivity measures

Using the above equations, we can derive connectivity measures from VAR models (see [8] for more details).

The cross-spectral density \( S(z) \) is obtained by

\[
S(z) = H(z) \Sigma_y H(z)^H. \tag{5}
\]

The COH is similar to \( S \), but normalized by the corresponding auto-spectra:

\[
\text{COH}_{ij}(z) = \frac{|S_{ij}(z)|}{\sqrt{S_{ii}(z)S_{jj}(z)}}. \tag{6}
\]

Variations of COH are imaginary coherence (iCOH) and partial coherence (pCOH):

\[
i\text{COH}_{ij}(z) = \frac{\Im(S_{ij}(z))}{\sqrt{S_{ii}(z)S_{jj}(z)}} \tag{7}
\]

\[
p\text{COH}_{ij}(z) = \frac{|G_{ij}(z)|}{\sqrt{G_{ii}(z)G_{jj}(z)}} \tag{8}
\]

with

\[
G_{ij}(z) = A(z)^H \Sigma_z^{-1} A(z). \tag{9}
\]

While the above measures are symmetric, the PDC provides directional connectivity information:

\[
\text{PDC}_{ij}(z) = \frac{|A_{ij}(z)|}{\sqrt{A_{ij}(z)A_{ji}(z)}}. \tag{10}
\]

The partial directed coherence factor (PDCF) incorporates the residual error covariance in the normalization:

\[
\text{PDCF}_{ij}(z) = \frac{|A_{ij}(z)|}{\sqrt{A_{ij}(z)\Sigma_z^{-1} A_{ij}(z)}}. \tag{11}
\]

The generalized partial directed coherence (GPDC) is a generalization of the PDC that does not depend on signal scale:

\[
\text{GPDC}_{ij}(z) = \frac{|A_{ij}(z)|}{\sigma \sqrt{A_{ij}(z)\text{diag}(\Sigma_z)^{-1} A_{ij}(z)}}. \tag{12}
\]

Another measure that provides directed connectivity information is the DTF:

\[
\text{DTF}_{ij}(z) = \frac{|H_{ij}(z)|}{\sqrt{H_{ii}(z)H_{jj}(z)}}. \tag{13}
\]

The full-frequency DTF (ffDTF) is similar to the DTF, except that normalization is performed over the full frequency range:

\[
\text{ffDTF}_{ij}(z) = \frac{|H_{ij}(z)|}{\sqrt{\sum_i H_{ii}(z)H_{jj}(z)}}. \tag{14}
\]
While the fDTF reveals direct and indirect connections, the direct DTF (dDTF) only shows direct connections:
\[ \text{dDTF}_{ij}(z) = p\text{COH}_{ij}(z) \cdot \text{fDTF}_{ij}(z). \]

The directed coherence (DCOH) can be understood as a generalization of the DTF, just like GPDC is a generalization of the PDC:
\[ \text{DCOH}_{ij}(z) = \frac{\sigma_j |H_{ij}(z)|}{\sqrt{H_i(z) \text{diag}(\Sigma_i) H_i^H(z)}}. \]

### 2.5. BCI simulation

Application of connectivity measures in BCIs based on MI requires estimation of these measures from a single realization of the time series. More channels and higher frequency resolution (model order) require longer estimation windows. However, we cannot increase the estimation window length indefinitely, because the time series are required to be stationary throughout the window. In addition, long estimation windows lead to a slow BCI response.

We propose a procedure that solves the problem of single-trial connectivity estimation in two steps. First, we compute connectivity estimates with low time and frequency resolution from all 45 available sources with estimation windows of length 4.5 s. Sources are ranked based on these estimates, and eight sources are selected for further use (a detailed description of source ranking and selection follows below). Second, the final connectivity estimates are obtained from the subset of eight sources with windows of length 1.5 s. In both steps, optimal model orders are determined by Akaike’s final prediction error (FPE) [33].

In a typical BCI experiment, the first step (initialization) is performed on training data recorded without feedback. Then, single-trial connectivity estimation and classification can be performed on the ongoing EEG. We simulate this strategy on pre-recorded data with an outer cross-validation loop. Classifier training requires a further nested cross-validation loop to optimize the time of training. Figure 3 shows schematically the functional blocks of our procedure. In the following sections, we will describe each block in detail.

#### 2.5.1. Outer cross-validation.

We employ a cross-validation scheme for validating the BCI system. From six available runs in a recording session, one is used for testing and the five remaining runs for training. This is repeated until every run was in the testing set once. Thus, trials from the same run can never occur in both the training and the testing set, and every trial is used for testing exactly once (see also [34]). The training sets are used for initialization of our procedure, and results are obtained from the respective testing sets.

#### 2.5.2. Initialization.

First, the unmixing matrix for extracting source signals from the EEG is estimated. This is accomplished by applying Infomax ICA [29] to the training data set. The unmixing matrix serves as a spatial filter that extracts the source signals from the EEG.

Next, the connectivity estimator fits VAR models to a sliding window of length 4.5 s over all 45 source signals, and subsequently extracts connectivity measures. The window is moved in steps of 0.3 s from \( t = 0 \) (cue) to \( t = 6 \) (end of trial), creating 21 time segments. This results in a four-dimensional data set \( c_k(i, j, f, t) \) for each trial \( k \), which describes a measure of connectivity from channel \( j \) to channel \( i \) at frequency \( f \) and time \( t \) relative to the cue. BP is estimated from the source autospectra obtained by fast Fourier transform. Although BP is not a connectivity measure, it can be treated as \( c_k(i, j, f, t) = 0 \)
for \( i \neq j \). We extract connectivity measures in the frequency range from 0–50 Hz in steps of 0.5 Hz. Once all connectivity measures are extracted, the following steps are performed for each measure separately.

The class correlator estimates the correlation coefficient \( r(i, j, f, t) \) of \( c_k \) with corresponding class labels \( l_k \). A class label is a numeric value that indicates if a trial contains hand MI (\( l_k = 1 \)) or foot MI (\( l_k = -1 \)). Thus, \( r \) indicates how strongly a connectivity measure can distinguish between both types of MI in channel pair \((i, j)\) at time \( t \) and frequency \( f \). A high positive correlation relates to MI of the same type. A high negative correlation relates to foot MI.

With the information obtained so far, the eight most important sources are selected. Sources that form a network with connections that differentiate strongly between classes are considered the most important. Thus, sources are ranked by the amount of correlation found in their in- and out-going connections, according to the ranking criterion \( e \) defined in (17).

\[
e_m = \sum_{f} \sum_{t} \left( \sum_{j} r^2(i, m, f, t) + \sum_{j} r^2(m, j, f, t) \right) . \tag{17}
\]

Once a subset of sources is selected, the unmixing matrix is pruned to only yield the selected sources. Connectivity estimates are obtained for these sources in a sliding window of length 1.5 s and the class-correlation \( r(i, j, f, t) \) is obtained for the estimates.

\( p(i, j, f, t) \)-values are obtained from the asymptotic normal distribution of \( r \), for testing against the null-hypothesis of no correlation. To determine which correlations are statistically significant, the significance mapper attempts to control the false discovery rate (FDR) [35] so that the probability of falsely detected correlations does not exceed 0.01. To reduce noise in the significance maps, we modified the FDR method so that correlations are only considered significant if they form a cluster of minimum size in the \( i/j/f/t \)-plane. Minimum cluster size was set to an area of 2 s Hz.

Features for classification are extracted from frequency bands where the connectivity exhibits significant class-correlation. The band selector identifies a list of frequency bands and connections \((i, j)\) where \( r \) is significant for a duration of at least 2 s. For functional connectivity measures, where \( c_k \) is symmetric, only one of the equivalent connections \((j, i)\) and \((i, j)\) is used. If no bands are found, the system resorts to a broad default band from 8 to 28 Hz for each connection.

A shrinkage linear discriminant analysis classifier [36] is trained on the features as follows. We use a nested cross-validation loop to estimate the classification performance for each time segment. In a leave-one-out procedure, each trial of the original training set is used for validation exactly once. A margin of ten trials before and after the validation trial is excluded from classifier training. This is repeated for each time segment. The final classifier is then trained on the best performing segment using all trials from the original training set.

- **Table 1.** ANOVA results for classification performance. The factors Selection, Features and Session are abbreviated as \( f_1 \), \( f_2 \), and \( f_3 \), respectively. \( F \) is the value of the \( f \)-statistic, with degrees of freedom \( \text{DFn} \) and \( \text{DFd} \).

| Effect       | DFn | DFd | \( F \)  | \( p \) | Sig |
|--------------|-----|-----|---------|--------|-----|
| \( f_1 \)    | 11  | 143 | 2.321   | 0.012  | *   |
| \( f_2 \)    | 11  | 143 | 28.588  | 0.000  | **  |
| \( f_3 \)    | 1   | 13  | 5.640   | 0.034  | *   |
| \( f_1:f_2 \)| 121 | 1573| 1.418   | 0.003  | **  |
| \( f_1:f_3 \)| 11  | 143 | 0.632   | 0.799  |     |
| \( f_2:f_3 \)| 11  | 143 | 0.678   | 0.758  |     |
| \( f_1:f_2:f_3 \)| 121 | 1573| 0.731   | 0.986  |     |

2.5.3. Single-trial connectivity classification. Single-trial analysis is performed on the testing set. The testing set was not used during initialization, thus this analysis is performed on completely unseen data. First, the pruned unmixing matrix obtained during initialization is used to extract selected sources from the testing set EEG. Connectivity measures are obtained from a sliding window of length 1.5 s, and selected frequency bands are extracted as features. Subsequently, the features are classified by the classifier trained during initialization.

We measure classification performance with Cohen's kappa \( \kappa \). To obtain a simple and robust measure of performance, we estimate \( \kappa(t) \) for each time segment between 1 s and 4 s after the cue, and report the median of \( \kappa(t) \) from these segments as classification performance.

2.6. Data analysis

Each connectivity measure was independently applied to source channel selection and classification. We analyzed classification performance across subjects and factors with repeated measures analysis of variance (ANOVA). The independent variable was \( f_1 \), and the three factors were \( f_1 \) (Selection, 11 levels), \( f_2 \) (Features, 11 levels), and \( f_3 \) (Session, 2 levels). Significant factors were further analyzed with Holm–Bonferroni corrected paired \( t \)-tests.

### 3. Results

Table 1 shows the results of the ANOVA for classification performance. There are highly significant \((p < 0.01)\) effects for \( f_2 \) (Features) and the interaction between \( f_1 \) (Selection) and \( f_2 \) (Features). The factors \( f_1 \) (Selection) and \( f_3 \) (Session) were significant \((p < 0.05)\).

Post tests were performed for significant effects. Post tests of the interaction between \( f_1 \) (Selection) and \( f_2 \) (Features) led to similar results as analyzing the main effects; we could not identify an interpretable pattern. Thus, only the main effects are presented for improved clarity. Post tests found no significant differences in \( f_1 \) (Selection) (see figure 4). The results of post tests on factor \( f_2 \) (Features) are shown in figure 5. BP performed best for classification. However, it was not significantly better than S, fIDTF and dDFT. All other measures performed significantly worse. Factor \( f_3 \) (Session) is shown in figure 6. Classification accuracy in the second session was significantly better than in the first session.
Figure 4. Factor $f_1$ (Selection). Left: classification performance $\kappa$ (mean and standard deviation) for each measure used for source channel selection. Right: significant (gray) and highly significant (black) differences between measures. Note. Since there are no significant differences, the plot is white and does not contain gray or black boxes.

Figure 5. Factor $f_2$ (Features). Left: classification performance $\kappa$ (mean and standard deviation) for each measure used as feature for classification. Right: significant (gray) and highly significant (black) differences between measures.

Figure 6. Factor $f_3$ (Session). Left: classification performance (mean and standard deviation) for each session. Right: significant (gray) differences between sessions.

The optimal VAR model order, which depends on the data, was different every time the VAR model was fitted. Table 2 shows the relative amount of selected model orders. For 45 sources, a model order of $p = 2$ was chosen most frequently. For the reduced set of eight sources, a model order of $p = 4$ was selected in the majority of cases.
Table 2. Optimal model orders determined by Akaike’s FPE for VAR models fitted to 45 and 8 sources, respectively. \( p \) is the model order, and fraction is the total percentage of each model order being selected.

| \( p \) | 45 sources | Fraction | 8 sources | Fraction |
|---|---|---|---|---|
| 1 | 22% | 4 | 86% |
| 2 | 77% | 5 | 8% |
| 3 | 1% | 6 | 3% |
| 4 | <1% | 7 | 2% |
| | | 8 | 1% |

4. Discussion

We proposed a procedure for an offline simulation of single-trial connectivity BCIs. This procedure can be easily applied in an online BCI. Currently, we are working toward implementing an online version of that approach. There is no outer cross-validation in an online setting. Instead, the training set is replaced by pre-recorded screening data, and the testing set is replaced by the online data stream as it is recorded.

We selected the optimal time to train the classifier based on estimates of classification performance obtained with the inner cross-validation loop. Ideally, the feature selection would also utilize classification performance as a ranking criterion. However, this would require even more cross-validation runs, thus increasing the initialization time dramatically. With an online BCI application in mind, we chose the computationally much less demanding feature selection based on correlation.

We validated the procedure by obtaining BP classification performance similar to other studies (e.g. [37, 16]). Also, the finding that classification performance improved in the second session is plausible, especially since only naive subjects participated in our measurements.

In a recent study, S, COH, and phase coupling were shown to yield similar results in terms of classification performance [16]. The authors concluded that this was caused by a bias toward zero phase between EEG channels. Indeed, our results support this conclusion. With ICA preprocessing, we effectively reduced zero-phase bias, leading to significantly worse classification with COH features. Thus, it is reasonable to assume that the COH between sources includes less task-relevant information than spectral power measures such as BP. In contrast, COH between EEG channels contains similar information as spectral power measures due to highly correlated EEG channels.

Our results indicate that features derived from univariate BP perform at least as well as any connectivity measure. Clearly, measures such as COH as well as the unmodified DTF are not suitable for a BCI application. However, S, dDTF, and fDTF perform similarly to BP. For S, this is not surprising, since the cross-spectral density is a generalization of the spectral density, the basis of BP.

The dDTF is similar to the fDTF, but only characterizes direct connections [38]. The fact that they perform equally well indicates that this difference might not be relevant for discriminating MI tasks.

ICA is ambiguous in terms of component order and scale. VAR models are invariant to permutations of the signals, but signal scale has a direct impact on the relative scaling of the VAR coefficients. This might introduce an error in the magnitude of connectivity measures. However, this error can be mitigated by using scale invariant connectivity measures such as GPDC or DCOH. We use both scale dependent and scale invariant measures, which allows us to determine the practical relevance of scale ambiguity for MI classification. GPDC and DCOH are similar generalizations of the PDC and the DTF, respectively. This generalization makes the GPDC and the DCOH invariant to the scaling of the signals. However, the generalized measures do not outperform the unmodified measures in terms of classification performance. This indicates that MI classification is unaffected by the scale ambiguity of ICA sources.

Resolution of the proposed measures is limited due to the single-trial estimation constraint. However, a possible solution could be provided by employing regularization or sparse methods for VAR model fitting [39, 40]. By setting unimportant AR coefficients to zero, fewer free parameters need to be estimated. This, in turn, would permit higher model orders, more channels, or shorter estimation windows. Better time resolution would allow the BCI to react faster to the user’s input, and better frequency or spatial resolution might improve classification as more information about the underlying structures and processes is resolved. Furthermore, the implicit coefficient selection of sparse model fitting might remove the need for our procedure’s source selection pass.

5. Conclusions

Although there is potential for further improvement, we were able to show that reliable classification of MI tasks is possible with single-trial estimation of the effective connectivity measures fDTF and dDTF.

The procedure we proposed in this paper can be used to consistently evaluate connectivity measures for BCI use. Thus, it may provide the basis for further research in the direction of connectivity-based BCIs.

To validate our findings in an online BCI system, the described methods can be easily adapted to an online system, since no major changes in the signal processing work flow are required.

Acknowledgments

This work was supported by the FWF Project ‘Coupling Measures for BCIs’ (P20848-N15).

References

[1] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control Clin. Neurophysiol. **113** 767–91
[2] Ramoser H, Müller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement IEEE Trans. Rehabil. Eng. **8** 441–6
newborn EEG Neural Information Processing (Lecture Notes in Computer Science vol 7664) ed T Huang, Z Zeng, C Li and C Leung (Berlin: Springer) pp 683–91

[21] Erfaa S, Faes L, Tranquilli E, Orlic D and Nollo G 2009 Multivariate autoregressive model with instantaneous effects to improve brain connectivity estimation Int. J. Bioelectromagn. 11 74–79

[22] Faes L and Nollo G 2010 Extended causal modeling to assess Partial Directed Coherence in multiple time series with significant instantaneous interactions Biol. Cybern. 103 387–400

[23] Hyvarinen A, Zhang K, Shimizu S, Hoyer P O and Dayan P 2010 Estimation of a structural vector autoregression model using non-Gaussianity J Mach. Learn. Res. 11 1709–31

[24] Gomez-Herrero G, Atienza M, Egiazarian K and Cantero J L 2008 Measuring directional coupling between EEG sources NeuroImage 43 497–508

[25] Delorme A et al 2011 EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing Comput. Intell. Neurosci. 2011 1–12

[26] Manke S, Bell A J, Jung T P and Sejnowski T J 1996 Independent component analysis of electroencephalographic data Advances in Neural Information Processing Systems (Cambridge, MA: MIT Press) pp 145–51

[27] Bell A J and Sejnowski T J 1995 An information-maximization approach to blind separation and blind deconvolution Neural Comput. 7 1129–59

[28] Ding M, Bressler S L, Yang W and Liang H 2000 Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment Biol. Cybern. 83 35–45

[29] Hyvärinen A, Zhang K, Shimizu S, Hoyer P O and Dayan P 2010 Estimation of a structural vector autoregression model using non-Gaussianity J Mach. Learn. Res. 11 1709–31

[30] Faes L and Nollo G 2010 Extended causal modeling to assess Partial Directed Coherence in multiple time series with significant instantaneous interactions Biol. Cybern. 103 387–400

[31] Hyvärinen A, Zhang K, Shimizu S, Hoyer P O and Dayan P 2010 Estimation of a structural vector autoregression model using non-Gaussianity J Mach. Learn. Res. 11 1709–31

[32] Delorme A et al 2011 EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing Comput. Intell. Neurosci. 2011 1–12